



Escuela  
Politécnica  
Superior

# Desarrollo de un sistema de BI aplicado a la gestión deportiva



Grado en Ingeniería Informática

## Trabajo Fin de Grado

Autor:

Ana Lavallo López

Tutores:

Juan Carlos Trujillo Mondéjar

Alejandro Maté Morga

Junio 2016



Universitat d'Alacant  
Universidad de Alicante

## **Agradecimientos**

Este trabajo ha sido parcialmente financiado por el Ministerio español de Economía y Competitividad (MINECO / FEDER) bajo el marco del proyecto SEQUOIA-UA (Requisitos de gestión y metodología para el análisis de grandes volúmenes de datos) TIN2015-63502-C3-3-R.

Agradecer a mis tutores, Juan Carlos Trujillo Mondéjar y Alejandro Maté Morga su apoyo y guía en el desarrollo del proyecto, ha sido un verdadero placer contar con su ayuda durante la realización.

Por otro lado, también agradecer a los profesores Jose M. Chinchilla y Antonio Turpín, de la empresa Kinnetic, el apoyo proporcionado a la hora de interpretar datos deportivos y, definir indicadores. La empresa Kinnetic es una experta en la monitorización y seguimiento de deportistas de élite.

## **Dedicatoria**

Dedico mi trabajo a mi familia por su apoyo incondicional y a esa persona especial que llena mi vida de color.

# Índice de contenidos

<b>Prefacio</b> .....	5
<b>1. Business intelligence</b> .....	7
1.1    Sistemas OLAP.....	9
1.2    Almacenes de datos .....	15
1.3    Proceso ETL .....	17
1.4    Fuentes de datos.....	19
1.5    Herramientas de creación de informes.....	20
<b>2. Desarrollo del sistema</b> .....	23
2.1    Fuentes de datos.....	25
2.2    Diseño del almacén de datos.....	29
2.3    Definición del cubo.....	33
2.4    Proceso ETL .....	37
2.5    Publicación del cubo en Pentaho .....	39
2.6    Consultas.....	41
<b>3. Análisis</b> .....	42
3.1    Pentaho.....	42
3.2    Weka .....	47
3.3    Power Pivot.....	55
<b>4. Indicadores de rendimiento (KPIs)</b> .....	61
<b>5. Perspectivas futuras</b> .....	63
<b>6. Conclusiones</b> .....	65
<b>7. Bibliografía</b> .....	67

## Prefacio

---

La motivación para el desarrollo del presente proyecto viene suscitada por el contenido cursado en la asignatura optativa “Inteligencia de Negocio y Gestión de Procesos” de la rama de especialidad de sistemas de información perteneciente al cuarto curso del plan de estudios del Grado en Ingeniería Informática. Dicha asignatura es impartida por el departamento de Lenguajes y Sistemas Informáticos perteneciente a la Universidad de Alicante.

Este proyecto ha sido realizado bajo la supervisión, seguimiento y evaluación del Catedrático de Universidad D. Juan Carlos Trujillo Mondejar y Alejandro Maté Morga, Doctor en Informática por la Universidad de Alicante.

El objetivo fundamental del trabajo realizado consiste en la aplicación de Business Intelligence (BI) en un ámbito deportivo y así mejorar el proceso de entrenamiento de los equipos y los resultados de sus jugadores en distintas pruebas físicas mediante la identificación de los factores más importantes que afectan a su desempeño, así como la definición de indicadores clave de rendimiento (KPIs, *Key Performance Indicators*) y demostrar en un ejemplo la gran potencia de este conjunto de herramientas.

Para conseguir el objetivo del proyecto se han aplicado varias técnicas de Inteligencia de Negocio y Minería de Datos. En concreto, se ha llevado a cabo el diseño e implementación de un almacén de datos mediante modelado multidimensional, que integre toda la información relevante para el entrenamiento, desde la información física de los jugadores a su desempeño en las distintas pruebas. Una vez integrada la información, se han aplicado múltiples técnicas de minería, incluyendo técnicas de vector soporte y árboles de decisión, que han permitido identificar los factores físicos que resultan más importantes de cara a obtener mejores resultados en las pruebas.

Una dificultad añadida de este trabajo es que, para poder obtener indicadores de rendimiento adecuados de deportistas de élite, éstos se deben obtener integrando información recogida en los eventos deportivos en los que participen con la información biométrica a través de una serie de sensores y pruebas. Existen varios sitios webs donde se pueden descargar a través de APIs, indicadores de la participación de deportistas en

eventos deportivos. Sin embargo, las pruebas biométricas realizadas a deportistas son datos que no están en abierto ni se pueden obtener.

Por ello, en este trabajo hemos trabajado con datos reales a través de los índices recogidos en categorías de un club de fútbol de la Comunidad Valenciana. Este hecho, confiere un valor añadido al proyecto puesto que los indicadores obtenidos están basados en recogida de datos a través de sensores y otras pruebas realizadas. Aprovechamos el trabajo para agradecer la cesión de los datos y la colaboración mantenida con el club de fútbol.

# 1. Business intelligence

---

Con el paso del tiempo y la estandarización de las bases de datos cada vez se cuenta con una mayor cantidad de información la cual poder analizar y procesar para extraer aquellos datos que resulten relevantes a sus usuarios. Al igual que la cantidad de datos ha ido creciendo exponencialmente, la forma de almacenarlos y procesarlos también ha ido avanzando para facilitar su uso y manipulación, ya que no es lo mismo trabajar sobre cientos de datos, que sobre millones o billones como en la actualidad.

Hoy en día, y gracias a las tecnologías de las que disponemos, se obtiene grandes cantidades de información sobre todo lo que nos rodea. En las actividades diarias de cualquier organización, se genera gran cantidad de datos como resultado de las transacciones realizadas. Hay empresas que facturan grandes cantidades de ingresos exclusivamente con la venta de datos, ya sean hospitales vendiendo información sobre sus pacientes o empresas tecnológicas que rastrean tus movimientos dentro de un centro comercial (p.e. a que escaparates le has dedicado más tiempo). Millones de empresas necesitan analizar sus datos para facilitar la toma de decisiones estratégicas, por ejemplo, saber cuál ha sido su producto más vendido en una zona un determinado año o cual ha sido el corredor de mi equipo que más ha mejorado desde que implantamos el nuevo sistema de entrenamiento

Estas preguntas resultan muy difíciles de responder usando sistemas transaccionales tradicionales (OLTP), debido a que la información se encuentra dispersa en numerosas tablas y no están orientados al análisis. Debido a estos problemas, se introdujo el uso del business intelligence (inteligencia de negocio), donde el usuario no necesita ser un profesional de TI (tecnologías de la información) para utilizar dichos sistemas. Los sistemas de business intelligence integran la información de distintas fuentes de datos en un solo repositorio, optimizado para responder las preguntas estratégicas de la empresa.

El business intelligence (BI) se conoce como la habilidad para transformar los datos en información, y la información en conocimiento. Conseguido a través del almacenamiento y procesamiento de grandes cantidades de datos, de forma que se pueda analizar la información a través de un análisis sencillo.

### **Ventajas del uso de BI:**

- Automatiza la asimilación de la información, ya que la extracción y carga de los datos se realiza a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Los usuarios no dependen de informes programados, ellos mismos consultan la información que necesitan de manera dinámica.
- Posibilita la formulación y respuesta de preguntas clave para la organización.
- Permite identificar factores que causan el buen o mal funcionamiento de la empresa, así como detectar situaciones fuera de lo normal.
- Permite predecir el comportamiento del futuro, basándose en los datos históricos de la organización.
- Los usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva, no necesitan ser profesionales de TI.

Hay que resaltar que las técnicas de BI no son utilizadas exclusivamente por grandes medianas empresas, sino que pueden ser aplicadas por cualquier persona que necesite tomar decisiones informadas a través del análisis de sus datos. A continuación, ejemplificamos esta facilidad mediante dos ejemplos de aplicación de BI para dos casos notablemente distintos.

#### **Empresa de venta de productos:**

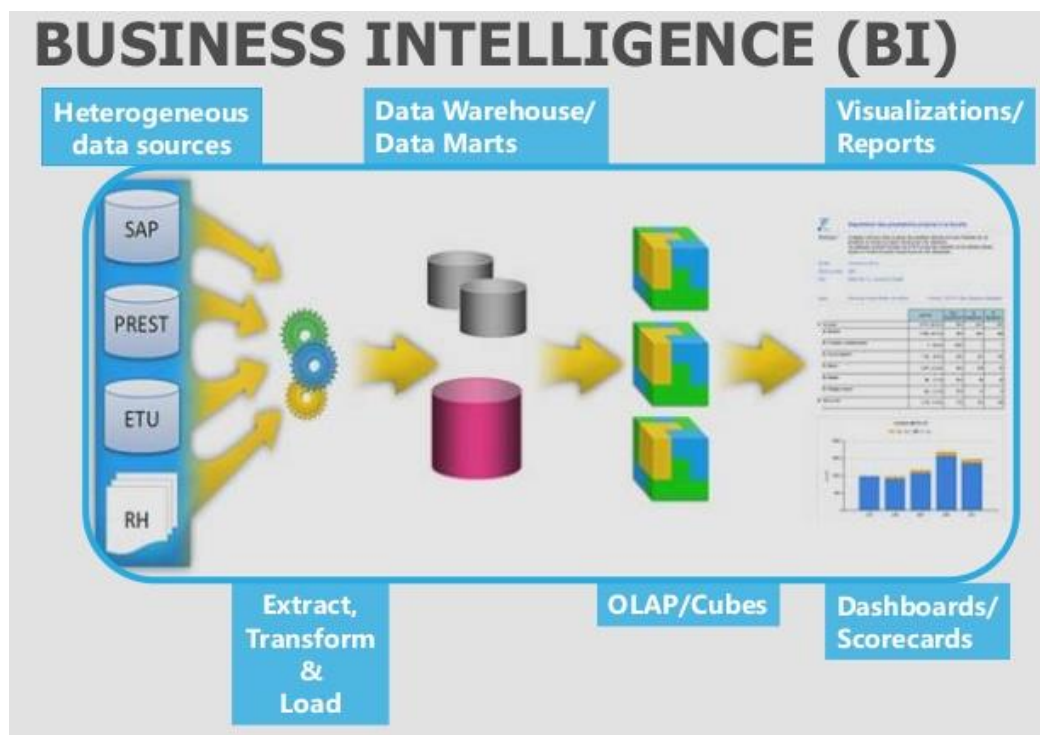
- *¿Quiénes son los mejores clientes?*
- *¿Cómo minimizar costos y maximizar las prestaciones?*
- *¿Cuál será el pronóstico de ventas del próximo mes?*

#### **Biblioteca:**

- *¿Cuál es la temática más consultada?*
- *¿Qué días hay mayor concurrencia, y por qué?*
- *¿Qué libros deben ser adquiridos?*



Para contestar a preguntas estratégicas como las mencionadas anteriormente, el BI utiliza un conjunto de metodologías, aplicaciones y tecnologías que permiten en tiempo real, el acceso interactivo, análisis y manipulación de información crítica para el usuario, dando así soporte a la toma de decisiones sobre el negocio (**Figura 1**). Los usuarios son capaces de analizar una gran cantidad de información de una manera muy sencilla y tomar las respectivas decisiones sin necesidad de ser expertos de las tecnologías de la información (TI).



**Figura 1.** Definición gráfica del concepto de Business Intelligence.

## 1.1 Sistemas OLAP

Antes de la aparición de los sistemas OLAP, se utilizaban exclusivamente sistemas transaccionales OLTP (OnLine Transaction Processing). Estos están orientados a las transacciones, por lo que soporta operaciones diarias de lectura, escritura, borrado y modificación de sus datos, mayoritariamente actuales. Son diseñados utilizando un modelo entidad-relación e implementados normalmente utilizando los motores típicos de bases datos (Oracle, SQL Server, Mysql...). Para utilizar estos sistemas es necesario ser profesional de las tecnologías de la información.

El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura concurrentes gracias a las propiedades ACID (Atomicity, Consistency, Isolation y Durability)

- Atomicidad asegura que la operación se ha realizado correcta y completamente.
- Consistencia asegura la integridad de la base de datos.
- Aislamiento asegura que dos transacciones que afectan a la misma información son independientes y no generan errores.
- Durabilidad asegura que una vez realizada la operación esta persistirá.

#### **Inconvenientes de los sistemas OLTP:**

- Inexistencia de la gestión de datos históricos.
- Normalmente se tienen distintas fuentes de datos (proveedores, clientes, componentes...)
- Sistemas difíciles de manejar por usuarios no expertos
- Rendimiento

El uso de estos sistemas requeriría integrar los datos, lo cual consumiría gran cantidad de tiempo y sería muy difícil de manejar por un analista sin grandes conocimientos sobre bases de datos.

Para solucionar estos problemas se introduce el business intelligence usando **sistemas OLAP** (On-Line Analytical Processing).

Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Esto implica el procesamiento de grandes volúmenes de datos históricos, procedentes de distintas fuentes para extraer información útil para el usuario.

Para representar la información, se utilizan los llamados cubos multidimensionales, que organizan la información alrededor de hechos (centro del análisis) y dimensiones (contexto del análisis). Sus elementos más importantes son:

- **Medidas:** Valores numéricos con los que se opera, pertenecientes a una tabla de hechos (nº clientes, importe de las ventas, nº de ventas...).
- **Dimensiones:** Conceptos que participan en el hecho, tales como clientes, almacenes, o jugadores, y que agrupan las medidas de forma que pueden ser

analizadas a distintos niveles de detalle, son almacenados en tablas de dimensiones.

- **Niveles:** Conceptos que participan en una misma dimensión con relaciones de jerarquía, y que permiten analizar las medidas de forma más agregada o detallada. Por ejemplo, provincia, comunidad autónoma y país serían niveles de la dimensión cliente.
- **Jerarquías:** Representa una relación de orden parcial entre dos o más niveles, y permiten navegar por los distintos niveles de una dimensión.

#### **Propiedades de las bases de datos OLAP:**

- Acceso a los datos suele ser solo de lectura, se dispone de valores agregados y resultados precalculados para permitir realizar las consultas más rápidamente.
- Organizadas según las necesidades analíticas, los datos están estructurados según las áreas de negocio.
- Asíncronas, no siempre se actualizan a tiempo real.
- Se suelen alimentar con información procedente de las bases de datos relacionales mediante un proceso de extracción, transformación y carga de los datos (ETL).
- Gestión de datos históricos a largo plazo, para realizar estudios de evolución a lo largo del tiempo, esto requiere que estas bases de datos mantengan un histórico a largo plazo, normalmente no inferior a cinco años.

En la **tabla 1**, se resumen la comparativa entre sistemas OLTP y OLAP:

**Tabla 1.** Principales diferencias entre los sistemas OLTP y OLAP.

	<b>OLTP</b>	<b>OLAP</b>
<b>Usuario</b>	Profesional de TI	Analista de información
<b>Objetivo</b>	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica
<b>Diseño de BD</b>	Orientada a aplicación (Basado en EE-R)	Orientado al tema (esquema estrella, copo de nieve)
<b>Tipo de datos</b>	Operacionales	Para la toma de decisiones
<b>Modelo de datos</b>	Normalizado	Desnormalizado
<b>Datos consultados</b>	Actuales	Actuales e históricos
<b>Tipos de consultas</b>	Repetitivas, simples	No previsibles, complejas
<b>Acciones</b>	Alta, baja, modificación y consulta	Carga y consulta
<b>Nº de transacciones</b>	Elevado	Medio – bajo
<b>Tamaño</b>	Pequeño - medio	Grande
<b>Referencia de tiempo</b>	Puede tener o no un elemento de tiempo	Tiene un elemento de tiempo
<b>Estructura</b>	Generalmente estable	Generalmente varía de acuerdo a su evolución y utilización

En los orígenes de la tecnología OLAP la mayor parte de las compañías asumió que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Sin embargo, otras compañías no tardaron en descubrir a través del uso de estructuras de bases de datos (esquemas de estrella y copos de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales para el OLAP.

A esta tecnología OLAP relacional se llamó ROLAP, en consecuencia, se adoptó el termino OLAP multidimensional (MOLAP). Las implementaciones MOLAP normalmente tienen mejor rendimiento y velocidad que la tecnología ROLAP, pero tienen

problemas de escalabilidad. Las implementaciones ROLAP son más escalables y normalmente más atractivas para los clientes ya que aprovechan las inversiones en tecnologías de bases de datos relacionales.

### **ROLAP (Relacional)**

Se ejecutan consultas SQL en bases de datos relacionales y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios. Accede directamente a los datos del data warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas.

### **MOLAP (Multidimensional)**

Una vez cargada la información en el sistema MOLAP, se realizan una serie de cálculos para obtener los datos agregados, a través de las dimensiones de negocio, rellenando la estructura de la base de datos multidimensional. Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas. Una vez este proceso está acabado, los usuarios solicitan informes a través de la interfaz y la lógica de la aplicación de base de datos multidimensional obtiene los datos.

Los datos están almacenados directamente en matrices y las operaciones de consulta están implementadas directamente sobre estas matrices (no están basados en SQL estándar)

A continuación, la **tabla 2**, compara las mejores y peores características de los sistemas descritos previamente.

**Tabla 2.** Ventajas (verde) versus inconvenientes (rojo) de los sistemas ROLAP Y MOLAP.

ROLAP	MOLAP
Brinda mucha flexibilidad, ya que los cubos son generados dinámicamente al momento de ejecutar las consultas.	Al realizar cambios sobre algún cubo, hay que recalcularlo totalmente para que se reflejen las modificaciones provocando una disminución importante de flexibilidad
Los datos de los cubos se calculan cada vez que se ejecuta una consulta sobre ellos, lo que provoca lentitud generando la respuesta de las consultas de los usuarios.	Las consultas son respondidas con mucha más rapidez ya los datos no son calculados en tiempo de ejecución.

Se crea una solución para cubrir las carencias de estos dos sistemas, un sistema denominado HOLAP.

### **HOLAP (Hibrido)**

Combina las arquitecturas ROLAP y MOLAP para crear una solución con las mejores características de cada una. Almacena agregaciones como un sistema MOLAP para mejorar la velocidad de las consultas y los datos se detallan en ROLAP para optimizar el tiempo en que se procesa el cubo.

## 1.2 Almacenes de datos

Los **Data Warehouse**, o almacenes de datos, son sistemas pensados para almacenar grandes volúmenes de datos sobre información histórica de las empresas, de forma que las consultas sobre la base de datos sean lo más eficientes posibles.

Uno de los primeros autores en escribir sobre los almacenes de datos fue Bill Inmon, definiéndolos a través de una serie de características:

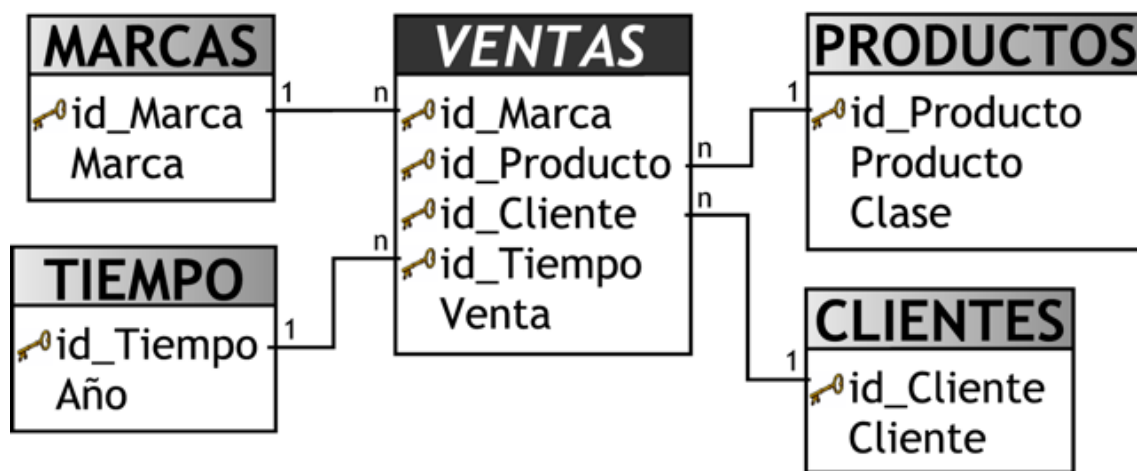
- **Orientados por temas:** En una tabla agrupamos los datos referentes al objeto de estudio. Aquí se encuentra la información que dará las respuestas a nuestras preguntas estratégicas.
- **Variables en el tiempo:** Se añade a los datos una referencia temporal para poder analizarlos en función del tiempo.
- **No volátiles:** La información no se modifica ni se elimina, una vez almacenado un dato, este se convierte en información de sólo lectura y se mantiene para futuras consultas.
- **Integrados:** Los datos están coherentemente agrupados a partir de las distintas fuentes de datos, dichos datos deben de ser consistentes

Los **Data Marts** se puede decir que son pequeños Data Warehouse centrados en un tema o un área de negocio específico dentro de una organización. Suelen contener más cantidad de información con más detalle. Supone buena opción para pequeñas y medianas empresas que no pueden afrontar el coste de poner en marcha un Data Warehouse. En la **tabla 3** se resumen las principales diferencias entre estos dos almacenes.

**Tabla 3.** Comparativa de Data Warehouse frente a los Data Marts.

<b>Data Warehouse</b>	<b>Data Marts</b>
<ul style="list-style-type: none"><li>• Información por temas de toda la organización.</li><li>• Complejo modelado de negocio.</li><li>• Se puede tardar años en construir e implementar.</li></ul>	<ul style="list-style-type: none"><li>• Información departamental (sub-temas).</li><li>• Más rápido de construir.</li><li>• La integración con DW puede ser compleja.</li></ul>

Para la implementación de estos sistemas se requiere, por una parte, de un sistema de gestión de base de datos como tecnología subyacente de almacenamiento, haciendo uso normalmente de un esquema estrella o de copos de nieve para su diseño, ya que optimiza las operaciones de consulta y reduce el uso de las operaciones de unión de las tablas (JOIN), la operación más costosa en una BD (**Figura 2**). Por otra parte, se necesita un conjunto de herramientas que permitan llevar a cabo el análisis de forma sencilla para que un analista de información que no sea profesional de TI sea capaz de analizar los datos.



**Figura 2.** Ejemplo de esquema en estrella.

La tabla central (ventas) se denomina tabla de hechos, contiene los hechos que serán utilizados por los analistas del negocio para apoyar la toma de decisiones, y referencia el número de dimensiones (las tablas que están alrededor). La clave primaria de la tabla de hechos está compuesta por las claves primarias de las tablas de dimensiones relacionadas con ella.

Los hechos son aquellos datos que residen en una tabla de hechos, pero no son claves primarias para referenciar las dimensiones. En este ejemplo el hecho es “Venta”. Los hechos son utilizados por los analistas de negocios para apoyar el proceso de toma de decisiones. Son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Contienen datos cuantitativos.

Usando estos métodos es posible producir informes que respondan preguntas estratégicas como, por ejemplo, cantidad de yogures de marca hacendado vendidos en 2015.



## 1.3 Proceso ETL

Mediante las herramientas y técnicas ETL (Extract, Transform, Load) se extraen los datos de las distintas fuentes, para cargarlos en nuestro almacén de datos (Data Warehouse o Data Mart). Para llevar a cabo este proceso, los datos han de ser limpiados, integrados y transformados para poder adaptarlos al diseño del almacén, ya que las fuentes de datos no siempre constan de datos homogéneos.

Para poblar el almacén con los datos obtenidos de las fuentes, las herramientas ETL realizan 3 etapas en este orden:

### EXTRACCIÓN

Esta primera etapa consiste en extraer los datos desde las fuentes de origen. Normalmente se fusionan los datos provenientes de diferentes sistemas, cada sistema puede usar una organización de los datos o formatos distintos. Los formatos de las fuentes suelen ser desde bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras. Dependiendo de la herramienta ETL empleada nos permitirá procesar distintos tipos de fuentes de datos.

### TRANSFORMACIÓN

Es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes.

Los casos más comunes en los que se deberán manipular los datos son:

- **Codificación.** Se encuentran distintas codificaciones para el mismo término. Por ejemplo, en el campo “Universidad” encontramos “Universidad de Alicante, U. Alicante y UA”
- **Medidas.** Encontraremos distintas unidades de medida para representar los atributos. Por ejemplo, el campo “Distancia” lo encontramos expresado en centímetros, metros y kilómetros según cada fuente.
- **Convenciones de nombramiento.** Un mismo atributo nombrado de diversas maneras. Por ejemplo, para referirse al nombre del proveedor encontramos, nombre, código y proveedor.

- **Fuentes múltiples.** Podemos encontrar la información de un campo representada en varias fuentes, se ha de escoger aquella que se considere más fiable y apropiada.
- **Valores nulos.** Al extraer los datos de las fuentes nos podemos encontrar con valores nulos para un campo.

En esta etapa también se realiza la limpieza de los datos, recolectando únicamente los datos relevantes para el estudio y asegurando que son de una buena calidad.

## **CARGA**

Por último, se realiza la carga de datos de los datos ya transformados, en el almacén.

En los casos en los que se tienen varios niveles de granularidad (jerarquías) se aplica el proceso de Rolling, en el que se almacena la información resumida a distintos niveles, correspondientes a las distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos las dimensiones almacenadas, por ejemplo, totales diarios, totales semanales, totales mensuales, etc)

La carga se podrá realizar en el momento actual o en cierto momento, de forma programada. Así podemos aprovechar los momentos en los que la base de datos soporte menos carga de usuarios para ejecutar el proceso ETL.

Existe gran variedad de herramientas ETL como, por ejemplo, Pentaho Data Integration (Kettle ETL), IBM Websphere DataStage, SAS ETL Studio, Oracle Warehouse Builder, BusinessObjects Data Integrator (BODI), Microsoft SQL Server Integration Services (SSIS) entre otras.

Para la realización de este proyecto se ha utilizado Pentaho Data Integration (Kettle ETL). Es una de las herramientas más conocida, por poseer versión de libre distribución y por sus extensas capacidades y facilidad de uso.

## 1.4 Fuentes de datos

Dependiendo de la herramienta ETL empleada nos permitirá procesar distintos tipos de fuentes de datos, con mayor o menor facilidad. Algunas de las fuentes más utilizadas en la actualidad son las siguientes:

**Bases de datos relacionales:** Nos resultará relativamente sencillo trabajar con ellas, ya que suelen ser altamente estructuradas. La mayoría de herramientas incluyen conectores con las principales de bases de datos.

**Bases de datos no relacionales:** (NoSQL, Analíticas, Hadoop...) Estas bases de datos surgen para el almacenamiento de grandes volúmenes de datos, su estructura es más sencilla que las tablas del modelo relacional. Buscan una eficiencia en la consulta de grandes volúmenes de datos y al contrario que el modelo relacional, no garantizan el modelo ACID (Atomicidad, Consistencia, Integridad y Durabilidad). Solo las herramientas más actualizadas los soportan.

**Archivos XML:** Lenguaje de marcas extensible, de naturaleza semiestructurada y empleado en numerosas aplicaciones, sobre todo las que tienen que ver en la transmisión de información a través de internet, aunque también se usa para el almacenamiento de información. Su naturaleza semiestructurada los hace más complejos de procesar, pero muchas de las herramientas ETL incorporan facilidades para procesarlos.

**Archivos JSON:** Java Script Object Notation, al igual que XML se trata de un lenguaje o sintaxis para el intercambio de información en forma textual y almacenamiento de la misma. Ocupa menos espacio que el XML y es más rápido y más fácil de procesar. Las herramientas ETL incorporan mecanismos para procesar este formato de datos.

**Archivos CSV:** Archivos de texto que permiten representar la información de una tabla. Contienen los valores separados por comas u otros separadores y cada fila representa una fila de la tabla. Tradicionalmente ha sido uno de los formatos más usados para la exportación e intercambio de datos desde la mayoría de aplicaciones como, por ejemplo: hojas de cálculo, aplicaciones de correo, bases de datos relacionales, etc. Prácticamente cualquier herramienta ETL nos permitirá trabajar con ellos.

**Archivos de hoja de cálculo:** Excel u otras hojas de cálculo, son una de las herramientas más frecuentemente usadas para procesar datos y analizarlos, aunque a un nivel mucho

más reducido que la tecnología de almacenes de datos. La mayoría de herramientas ETL los soportan y es fácil trabajar con ellos debido a su estructura de filas y celdas bastante parecida a una tabla de una base de datos relacional.

**Archivos de texto con contenido desestructurado:** Hay aplicaciones que almacenan o exportan sus datos en archivos de texto cuyo contenido está muy poco o nada estructurado. Como por ejemplo el log de un servidor web. Necesitaríamos una buena herramienta ETL que nos proporcione herramientas para trabajar con ellos, como por ejemplo el reconocimiento de expresiones regulares en el texto.

Cuanto más tipos de fuentes nos permita procesar la herramienta ETL mejor, pues mayor flexibilidad tendremos. Además, las herramientas ETL deben actualizarse y adaptarse a las nuevas fuentes que van apareciendo en el mercado.

Para nuestro proyecto hemos decidido utilizar la herramienta Pentaho Data Integration (Kettle ETL) ya que nuestras fuentes son archivos cvs y proporciona suficientes características para procesar este tipo de archivos.

## 1.5 Herramientas de creación de informes

Para una mayor facilidad a la hora de consultar los datos, los usuarios necesitan tenerlos agrupados y representados de una forma gráfica, para ello se utilizan herramientas de creación de informes, existen gran variedad de herramientas, algunas gratuitas y otras de pago. A continuación, vamos a nombrar alguna de ellas.

Pentaho contiene la herramienta **Pentaho Report Designer** (Figura 3). Es una herramienta gratuita que permite personalizar los informes, bien para ejecutarlos directamente o para publicarlos en una plataforma **Pentaho Report Designer** BI y que desde ahí puedan ser utilizados por los usuarios. Permite trabajar con múltiples orígenes de datos como JDBC, Olap4J, Pentaho Analysis, Pentaho Data Integration, XML en un entorno gráfico. Y permite la salida de los resultados en formatos como PDF, HTML, XLS, RTF y CSV.

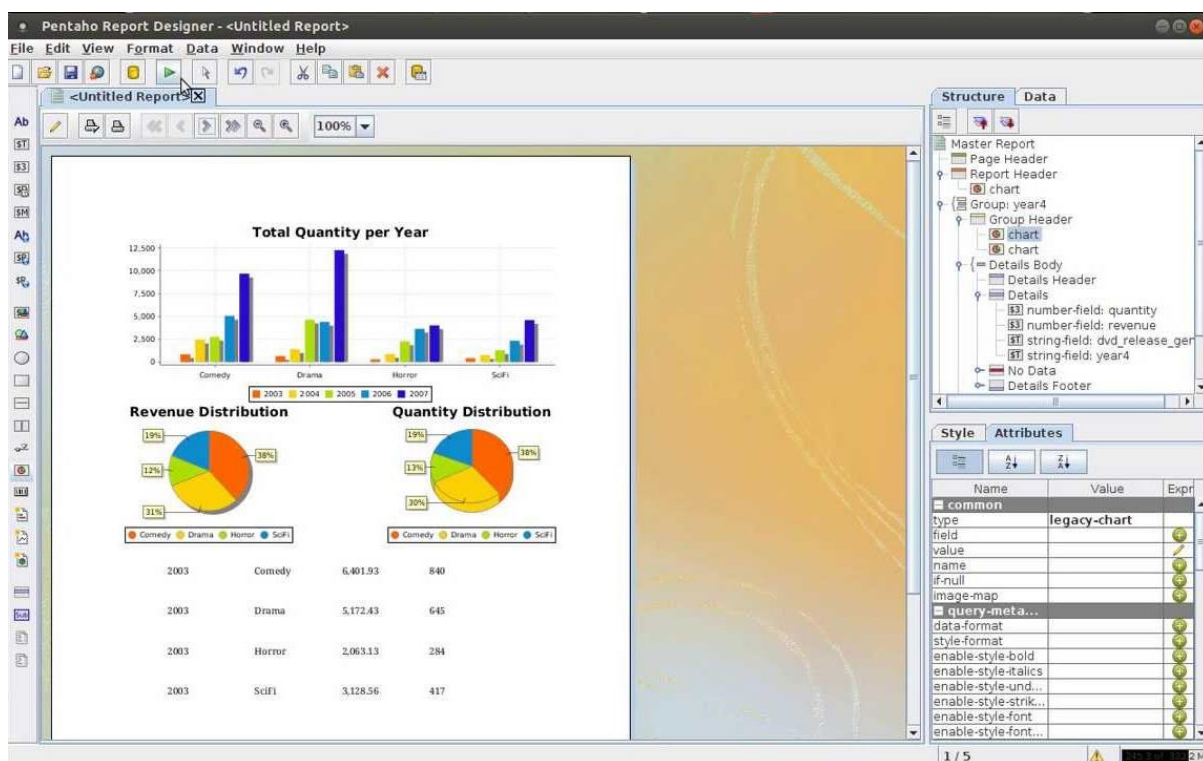


Figura 3. Ejemplo de la herramienta Pentaho Report Designer.

Microsoft ofrece la herramienta **Power BI** (Figura 4), es un conjunto de aplicaciones de análisis de negocios que permite analizar datos y compartir información. Dispone una versión gratuita y una de pago con más funcionalidades por 8,40€ por usuario por mes.

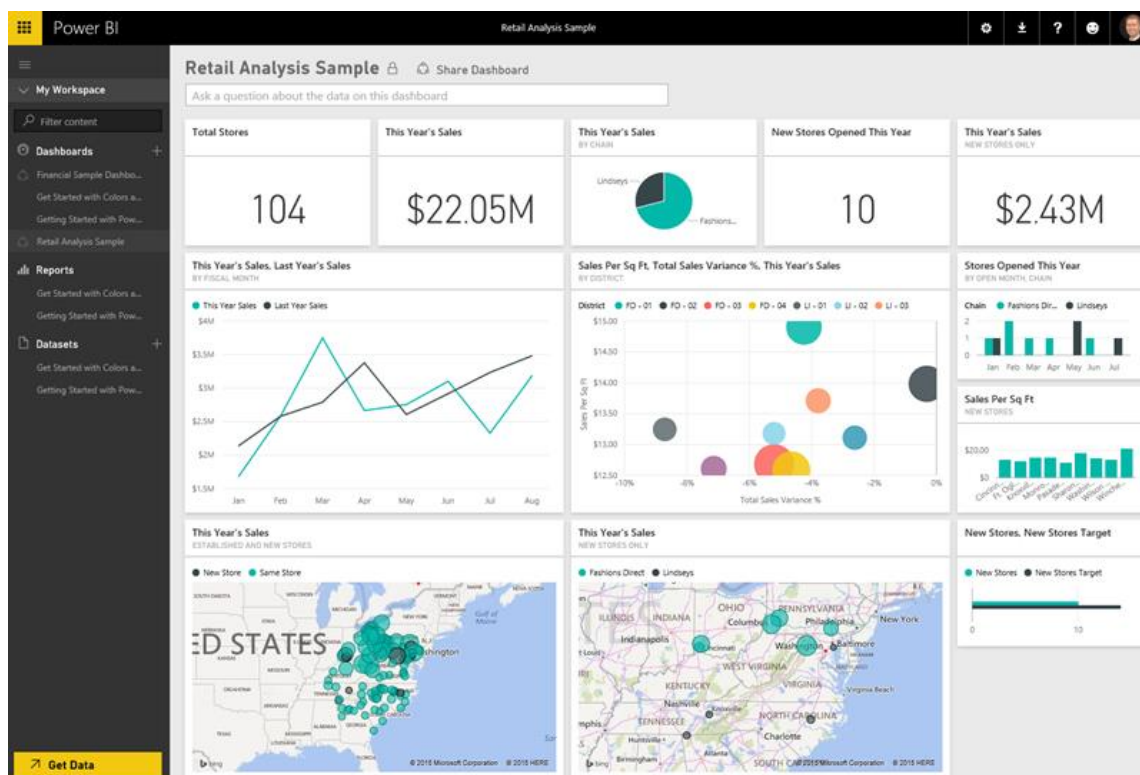


Figura 4. Ejemplo de la herramienta Power BI.

Otra herramienta que ofrece Microsoft es **Power Pivot** (Figura 5), es un complemento de Excel que se usa para realizar análisis de datos y crear modelos de datos sofisticados. Permite combinar grandes volúmenes de datos de orígenes diferentes, realizar análisis de la información rápidamente y compartir puntos de vista con facilidad. El uso de este complemento es gratuito siempre y cuando se tenga Excel, el coste de esta herramienta es de 135€.

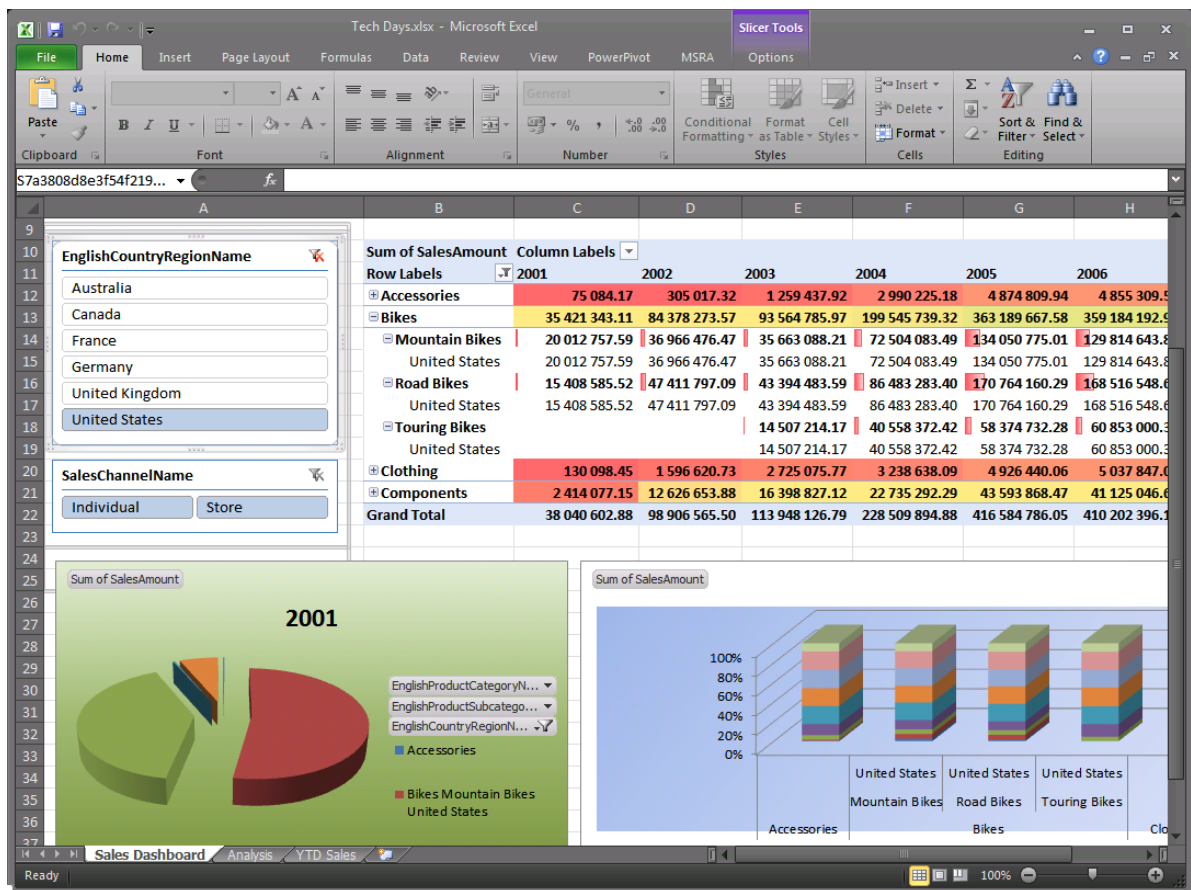


Figura 5. Ejemplo de la herramienta Power Pivot.

Esta ha sido la herramienta seleccionada para la creación de alguno de los informes de este proyecto ya que contaba con todas las características que se necesitaba y la Universidad de Alicante ofrece la posibilidad de obtener Excel de forma gratuita.

## 2. Desarrollo del sistema

---

Business intelligence es usado en muchos sectores de consumo y se ha convertido en una herramienta imprescindible para poder competir en los mismos. Las empresas de servicio disponen de la información de sus clientes y las empresas de producto son capaces de obtener dicha información y potenciar la relación con sus clientes. Pero este concepto aún no se ha aplicado sobre todo el potencial que ofrece el mundo deportivo. El mercado deportivo cuenta con millones de seguidores, eventos, jugadores, apuestas, etc. Se puede extraer mucha información de las personas que acuden a eventos deportivos, pueden ser analizados los comportamientos y hábitos de dichas personas, así como el consumo de otros productos y servicios en las instalaciones y sus alrededores. Se generan millones de datos cada día, y no están siendo aprovechados como se podría.

En el presente apartado se lleva a cabo la puesta en práctica de lo expuesto anteriormente, la introducción del business intelligence al mundo deportivo. En este caso se ha enfocado a la gestión de jugadores de un club de fútbol, así como la búsqueda de características en los jugadores que determinen mejor rendimiento.

En primer lugar, ha sido necesario obtener la información, tanto de los resultados de las pruebas de los jugadores como de sus aptitudes físicas. Para ello, se ha contado con la colaboración del Hércules de Alicante Club de Fútbol, quien ha proporcionado los datos de las pruebas que se han realizado a los jugadores de las distintas categorías del club obteniendo, datos generales, de bioimpedancia y datos resultantes de una batería de tests físicos como salto, flexibilidad y velocidad.

A continuación, esta información ha sido procesada e integrada en un data warehouse diseñado exclusivamente para este análisis. Para ello, se ha llevado a cabo un diseño multidimensional de los datos y, posteriormente, el diseño e implementación de los procesos ETL que han permitido cargar los datos obtenidos en el data warehouse.

Después, mediante diversas herramientas y técnicas de minería de datos, se han realizado análisis de los resultados obtenidos por los jugadores con el objetivo de la obtención de una o varias características físicas comunes en los jugadores que determinen los mejores resultados en las pruebas realizadas.

A raíz de estos análisis, se ha diseñado un cuadro de mandos que integre el conocimiento adquirido. Este cuadro de mandos representa las medidas e indicadores que, según el análisis realizado, se han mostrado determinantes para que los jugadores obtengan los mejores resultados. Además, este cuadro de mandos permite seguir el rendimiento de los jugadores de una forma dinámica.

Gracias a las medidas e indicadores obtenidos, los entrenadores pueden realizar un seguimiento más detallado de la evolución de los jugadores, permitiendo comparar el desarrollo de sus aptitudes físicas con los resultados obtenidos. De esta forma, pueden diseñar programas de entrenamiento a medida en función de los jugadores y los objetivos perseguidos.

Con la vista puesta en el futuro, se cuenta con la colaboración de la empresa KINETIC quien dispone de un software llamado SportsCode, el cual analiza de una forma muy eficaz videos deportivos. Este análisis de vídeos facilitará datos que permitirán ampliar el estudio aquí realizado y obtener un sistema de apoyo a la toma de decisiones para equipos deportivos aún más completo.



## 2.1 Fuentes de datos

Para la población de la base de datos se dispone de un documento Excel con los datos resultantes de los distintos tipos de prueba que se ha realizado sobre 90 jugadores. Estos datos han sido tomados en una sola medición en la temporada 2015/2016, incluyen 3 tipos de datos:

### **Generales (Figura 6):**

ID	EQUIPO	EDAD	FECHA NAC.	POSICION	ALTURA (cm)	Nº CALZADO	PESO (kg)	BMI	AGUA CORPORAL %	COMPLEXION FISICA	MASA OSEA (kg)	BMR (KCAL)	BMR (KJ)	EDAD METABOLICA	GRASA VISCERAL
1	alevin A	12	27-04-04	defensa	151	40	49,5	21,7	62,6	5	2,2	1304	5456	12	1
2	alevin A	12	27-04-04	centrocam	160	43	55,5	21,7	66,4	5	2,5	1511	6322	12	1
3	alevin B	10	26-8-05	portero	148	40	42	19,2	69,3	7	2	1194	4996	12	1
4	alevin B	11	23-4-05	defensa	137	36,5	40,9	21,8	65,4	5	1,8	1118	4678	12	1,5
5	alevin B	11	08-02-05	defensa	147	38,5	40,3	18,6	69,8	7	1,9	1158	4845	12	1
6	alevin B	10	18-08-05	centrocam	148	37,5	35,6	16,2	72,9	7	1,8	1072	4485	12	1
7	alevin B	10	19-06-05	centrocam	133	34	31	17,5	74,4	7	1,5	956	4000	12	1
8	alevin B	11	17-05-05	centrocam	146	40	41,1	19,3	72,7	8	2	1202	5029	12	1
9	alevin B	10	31-12-05	delantero	139	36	32,1	16,6	75	7	1,6	979	4096	12	1
10	alevin B	10	15-07-05	delantero	142	36	35,2	17,5	81,2	7	1,8	1075	4498	12	1

**Figura 6.** Datos generales de los jugadores.

Se ha mostrado como ejemplo 10 registros de los 90 disponibles. Los jugadores están divididos en 5 categorías de equipo agrupados por edades. En el 1er equipo se encuentran los jugadores de mayor edad y más profesionales, la siguiente categoría son los cadetes A y B, compuestos por jugadores de edades entre 14 y 16 años, la categoría infantil consta de jugadores de 13 y 14 años. La categoría alevín tiene 3 equipos A, B y C con jugadores de entre 10 y 12 años y por último los jugadores más pequeños se encuentran en la categoría benjamín, también formado por 3 equipos, el A, B y C y compuestos por jugadores de entre 8 y 10 años. Las categorías son divididas en equipos A, B y C dependiendo de la profesionalidad de los jugadores que lo componen, normalmente los mejores jugadores se encuentran en el equipo A.

### **Bioimpedancia:**

La bioimpedancia es una técnica que sirve para calcular el porcentaje masa grasa y masa muscular. La **Figura 7** representa el % de grasa que hay en las distintas zonas del cuerpo. Los niveles en los que son clasificados son generados por la máquina que realiza las pruebas para su agrupación.

GRASA CUERPO	NIVEL GRASA CUERPO	GRASA BRAZO DERECHO	NIVEL GRASA BRAZO D	GRASA BRAZO IZQUIERDO	NIVEL GRASA BRAZO I	GRASA TRONCO	NIVEL GRASA TRONCO	GRASA PIERNA DERECHA	NIVEL GRASA PIERNA D	GRASA PIERNA IZQUIERDA	NIVEL GRASA PIERNA I
15,4	6	13	4	14	6	21,3	4	8,2	2	7,3	2
10	6	9,5	3	10,2	3	13,1	2	5,4	2	6,1	2
9,4	3	8,5	3	8,3	3	12,2	2	6,1	2	6,1	2
15,7	6	12,5	3	13,7	4	21,9	4	8,1	2	9,1	2
9,3	3	7,7	3	9,5	3	14,3	3	4	2	2,1	2
6,7	3	4,9	2	6	2	10,7	2	2,6	1	1,4	1
9,1	3	8,5	3	7,7	3	13,8	3	4,5	1	3,5	1
6,2	3	5,4	2	7,3	2	10	2	1,6	1	1,5	1
9,1	3	7,1	2	7,5	2	14,5	3	3,6	2	3,2	2
5	3	2,8	2	2	1	9,4	2	1	1	1	1

**Figura 7.** Datos sobre el porcentaje y nivel de grasa.

En la **Figura 8** se ve representado los kilogramos de masa muscular en las diferentes zonas, al igual que en la tabla de la grasa los niveles son generados por la máquina.

MUSCULO CUERPO	NIVEL MUSCULO CUERPO	MUSCULO BRAZO DERECHO	NIVEL MUSCULO BRAZO D	MUSCULO BRAZO IZQUIERDO	NIVEL MUSCULO BRAZO I	MUSCULO TRONCO	NIVEL MUSCULO TRONCO	MUSCULO PIERNA DERECHA	NIVEL MUSCULO PIERNA D	MUSCULO PIERNA IZQUIERDA	NIVEL MUSCULO PIERNA I
39,7	4	2,1	3	1,9	3	19,2	3	8,4	6	8,1	6
47,5	5	2,6	4	2,5	4	23,9	4	9,5	7	9	6
36	3	1,9	3	1,8	3	17,7	3	7,5	5	7,1	5
32,7	4	1,7	4	1,5	3	15,1	3	7,4	7	7	7
34,6	3	1,9	3	1,6	3	16,4	2	7,5	5	7,2	5
31,4	2	1,7	3	1,5	2	14,8	1	6,9	5	6,5	4
26,7	3	1,4	3	1,2	2	11,8	1	6,3	6	6	5
36,5	4	52	4	1,8	3	17,3	3	8	7	7,4	5
27,6	2	1,5	3	1,2	2	12,4	1	6,4	5	6,1	5
31,7	3	1,8	3	1,7	3	13,8	1	7,5	7	6,9	5

**Figura 8.** Datos sobre los kilogramos y nivel de musculo.

### **Resultados de las pruebas:**

La **Figura 9** representa las distintas pruebas físicas que han sido realizadas a los jugadores. Estas pruebas están divididas en 3 categorías, salto, flexibilidad y velocidad.

Salto SJ	Salto CMJ	Salto CMJ+BRAZOS	% ELAST.	COORD.	FLEXIBILIDAD	TIEMPO 10M	TIEMPO 20M	AGILIDAD	TIEMPO 6X20 TOTAL	TIEMPO 6X20 MEDIA	% PERDIDA VELOCIDAD 6X20
15,1333	20,6333	23,5	36,34	13,89	24,5	2,04	3,59	16,15	22,4	3,73	8,53
10,98	14,1933	25,2466667	29,27	77,88	21,833333	2,02	3,56	15,8	21,65	3,61	9,01
15,5	15,08	23,57	-2,7312	56,31	30,166667	2,25	4,13	16,8867	24,59	4,10	5,09
22,9233	26,63	31,77	16,184	19,27	31	1,86	3,42	15,7667	23,37	3,90	14,22
16,0333	22,63	18,40	41,164	-18,70	21	1,96	3,54	15,67	23,49	3,92	11,86
18,3333	19,67	23,37	7,2727	18,81	15,5	2,04	3,65	17,82	24,05	4,01	4,93
20,6	21,43	24,37	4,0453	13,69	20,166667	1,97	3,75	15,96	24,69	4,12	8,86
22,1	25,07	27,57	13,424	9,97	32	2,09	3,72	18,36	24,22	4,04	9,10
14,9333	17,73	22,70	18,75	28,01	21,166667	2,05	3,63	16,45	24,06	4,01	8,38
22,7	20,60	30,80	-9,2511	49,51	22,833333	1,98	3,46	14,16	21,97	3,66	15,88

**Figura 9.** Datos resultantes de las pruebas físicas.

**Salto**, donde se encuentran las pruebas:

- Salto SJ (Squat Jum)



En esta prueba el jugador debe efectuar un salto vertical partiendo de la posición de media sentadilla (rodillas flexionadas a 90°) con el tronco erguido y las manos en la cintura. No se pueden realizar contra movimientos hacia abajo ni ayudarse con los brazos. En esta prueba se analiza la fuerza explosiva del jugador. Se mide la altura alcanzada en el salto expresado en centímetros.

- Salto CMJ (Counter Movement Jump)



En este caso el jugador inicia el movimiento desde una posición erguida para luego bajar y volver a subir lo más rápido posible al salto. Sigue con las manos en la cintura y no puede ayudarse con ellas. Se analiza el contramovimiento y la elasticidad. También se mide la altura alcanzada en el salto expresada en centímetros.

- Salto CMJ + Brazos

Se realiza el salto CMJ pero en esta prueba se cuenta con la ayuda de los brazos para impulsarse. Al igual que los demás saltos se mide la altura alcanzada y el resultado está expresado en centímetros.

Mediante la diferencia de resultado entre el salto SJ y el CMJ se calcula el porcentaje de elasticidad del jugador. Para la realización del análisis se ha basado en tablas de normalidad que están validadas para determinar si los datos obtenidos se encuentran en valores de normalidad o no. En este caso para que la elasticidad se encuentre en valores de normalidad debe estar por encima del 10%.

Con la diferencia del salto CMJ y el CMJ + brazos se calcula la coordinación. Al igual que para la elasticidad hay que basarse en tablas de normalidad, que para encontrarse en valores de normalidad la coordinación debe estar por encima del 12%.

**Flexibilidad**, consiste en medir la flexibilidad de la cadena posterior, la parte de los isquios. Los jugadores se sientan en el suelo y apoyan la planta de los pies en el aparato para luego realizar una flexión de cadera empujando una pieza hasta que no puedan más, se anota la medida expresada en centímetros. Con esta medida dependiendo de la edad y del género se determina el percentil en el que está basándose en unas tablas para poder determinar si tiene que mejorar esta cualidad o no.

**Velocidad** donde se tienen las pruebas:

- 10 Metros



Son utilizadas células fotoeléctricas como las de la imagen para medir el tiempo en recorrer 10 metros. Ya que los tiempos son muy pequeños y suele haber escasa diferencia entre los jugadores. Es medido en segundos.

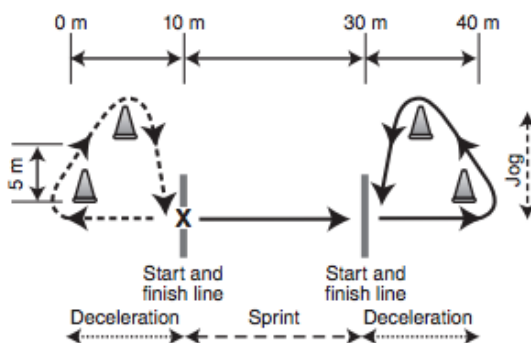
- 20 Metros

Se realiza igual que la prueba de 10 metros, pero con doble distancia.

- Agilidad

Se realiza un circuito de agilidad y el resultado es el tiempo que se ha tardado en realizar. Esta expresado en segundos.

- 6 series de 20 metros



Se realiza 6 veces este circuito de velocidad. De esta última prueba obtenemos tiempo total de realización, el tiempo medio de las 6 series y el porcentaje de velocidad perdido a lo largo de la prueba.

## 2.2 Diseño del almacén de datos

El diseño del almacén de datos es una de las piezas más importantes del desarrollo, es la base desde donde se construye todo el proyecto, por lo que es muy importante verificar que su diseño es el correcto, ya que una vez avanzada la construcción, volver atrás para corregir errores es una tarea complicada.

Se ha elegido un modelo estrella para su realización, ya que no es una base de datos excesivamente grande, tiene una baja dimensionalidad y frente a un modelo de copo de nieve este ofrece un mayor rendimiento a la hora de procesar.

Mediante el análisis de las fuentes de datos ha sido diseñado el siguiente almacén (**Figura 10**). El esquema estrella se caracteriza por tener una tabla de hechos central, en este caso sería la tabla “Resultados”, esta está formada por las claves primarias de las tablas de dimensiones y las medidas del cubo (atributo resultado). Por otra parte, se han creado 14 tablas de dimensiones, una con los datos generales de los jugadores, otra los tipos de prueba y sus categorías y por último una tabla por cada una de las mediciones de bioimpedancia. Este diseño refleja los datos de los jugadores en una sola medición en una temporada, la del 2015/2016. En el caso de tener más mediciones, ya puede ser en la misma temporada u otras distintas, será añadida una dimensión tiempo para poder almacenar los datos agrupados en los distintos espacios temporales.

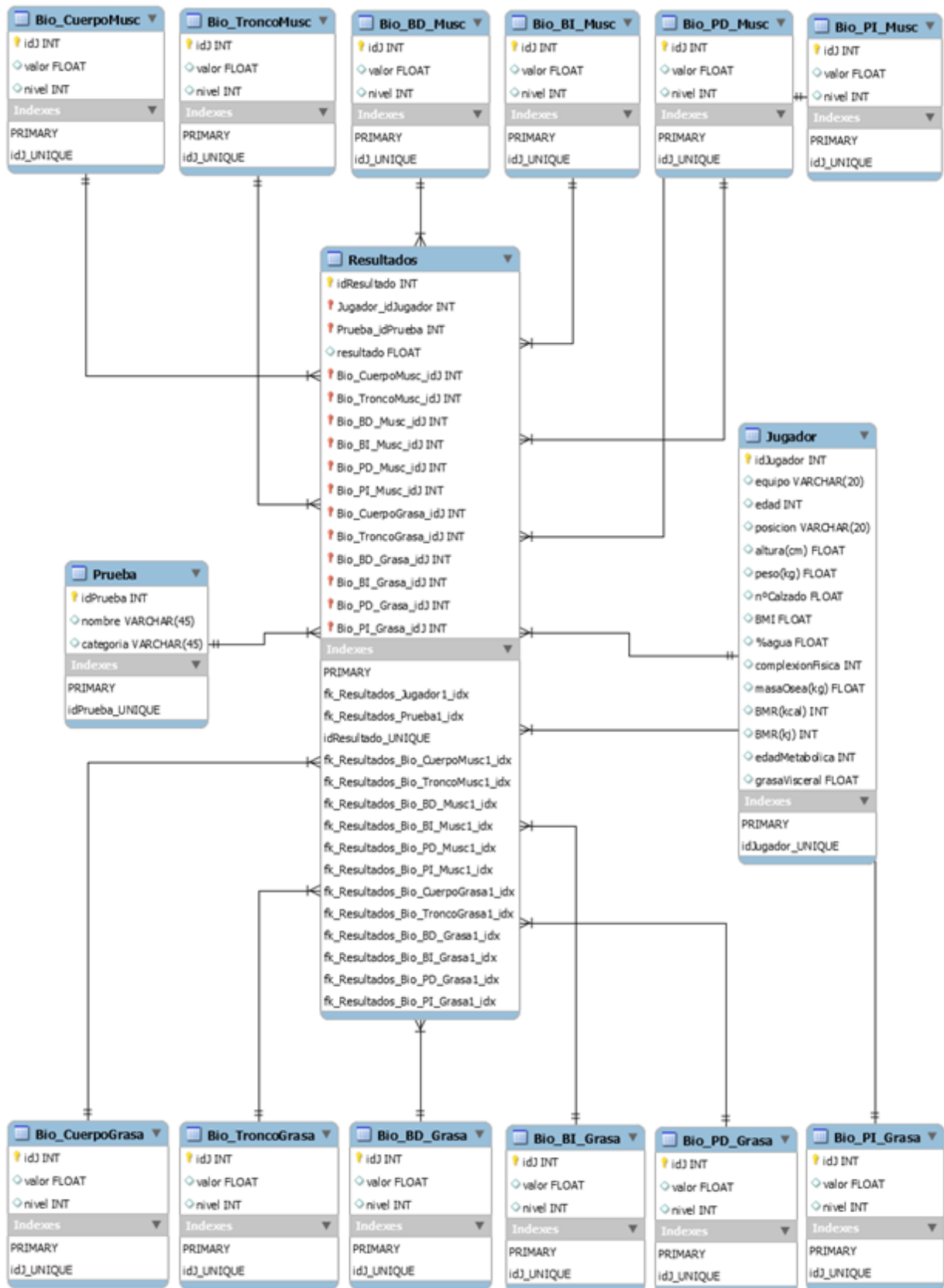
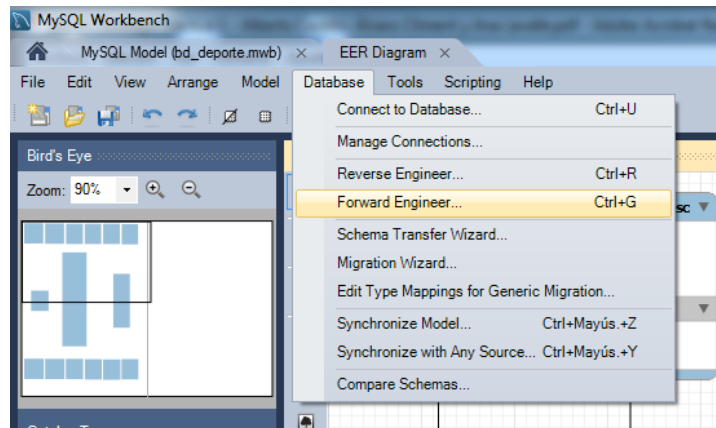


Figura 10. Diseño del almacén de datos.

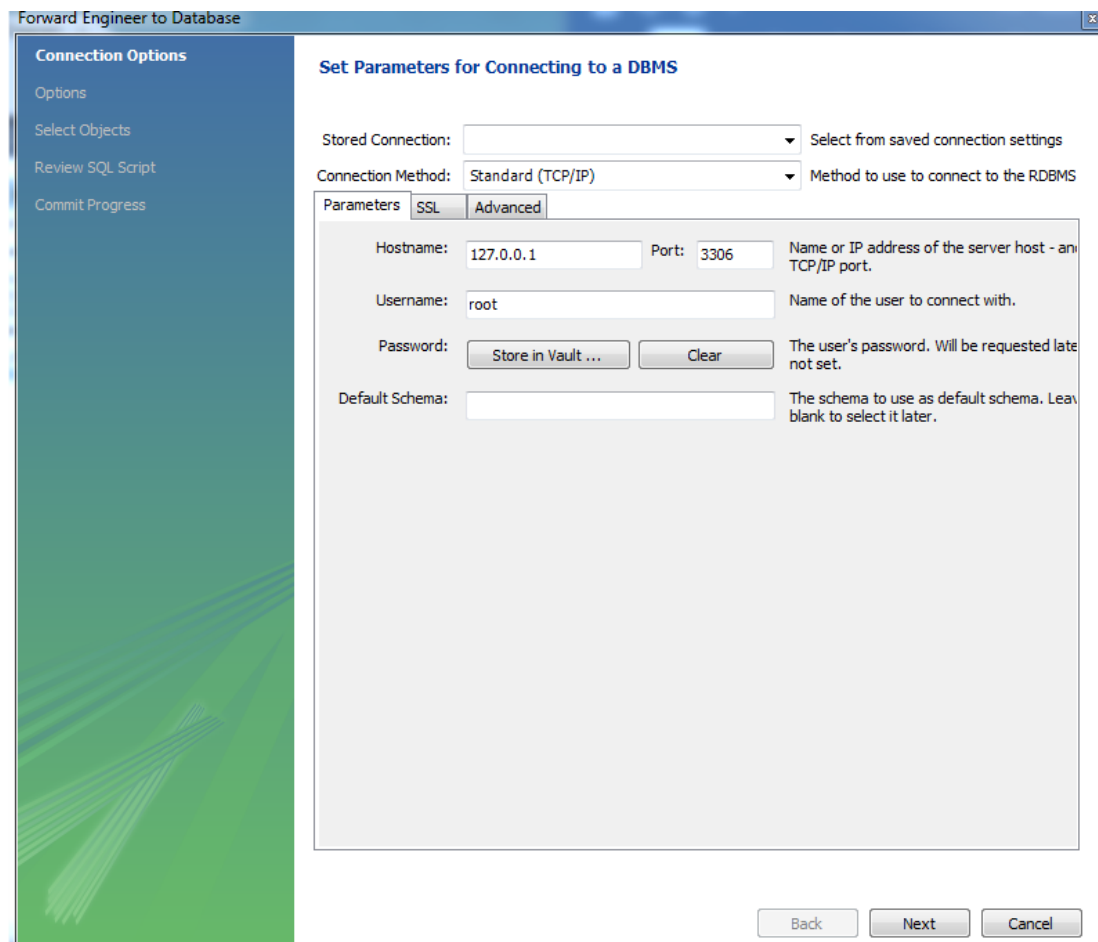
Una vez revisado y aceptado el diseño, mediante la herramienta MySQL Workbench se procede a pasar el diseño a MySQL. Estos son los pasos más relevantes que se han tenido que realizar:

Se selecciona la opción “Forward Engineer” (**Figura 11**).



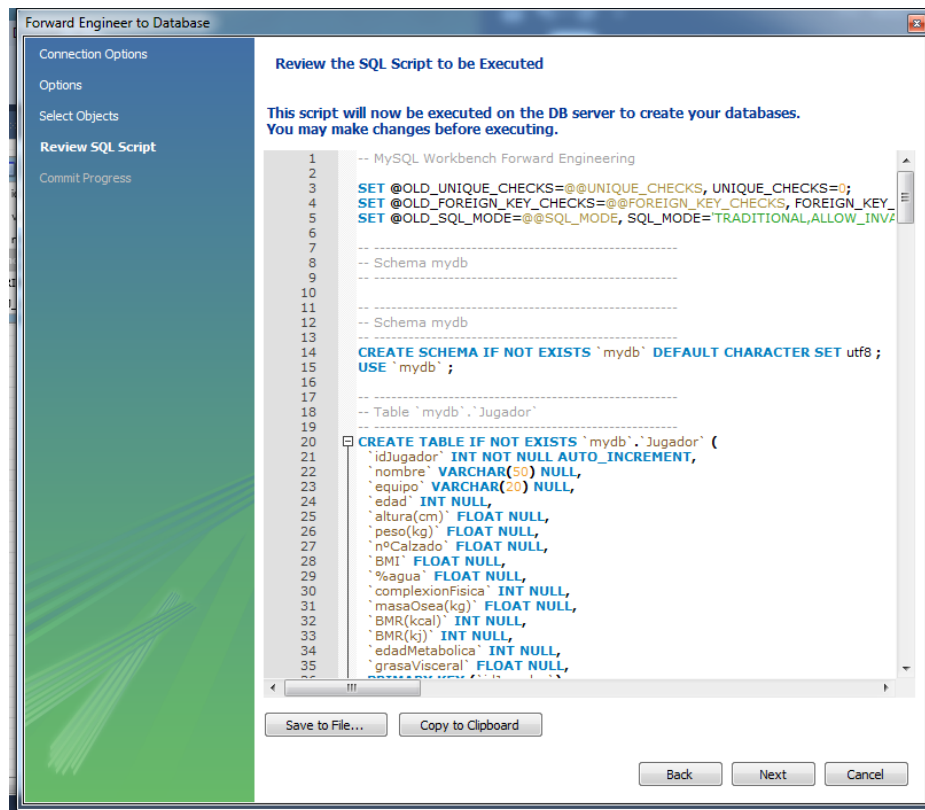
**Figura 11.** Selección de la opción “Forward Engineer en la herramienta MySQL Workbench.

Son indicados los parámetros de conexión (**Figura 12**).



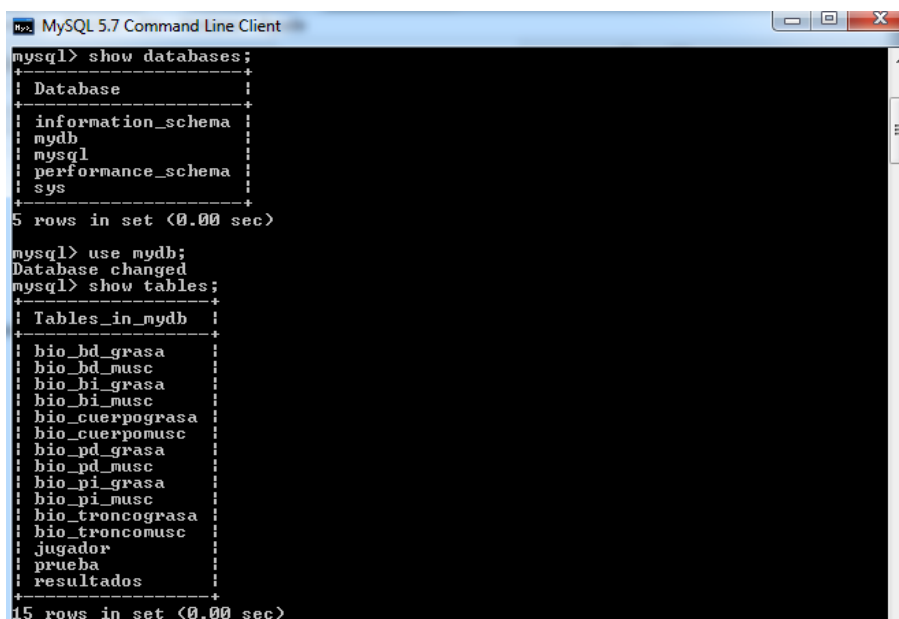
**Figura 12.** Parámetros de conexión con la base de datos.

En este último paso el programa muestra el script de SQL que va a ser ejecutado para crear la base de datos del diseño que había sido realizado (**Figura 13**).



**Figura 13.** Script de la creación de la BD.

Una vez creada la base de datos y su conexión se puede ejecutar MySQL para comprobar que la base de datos has sido creada correctamente (**Figura 14**).



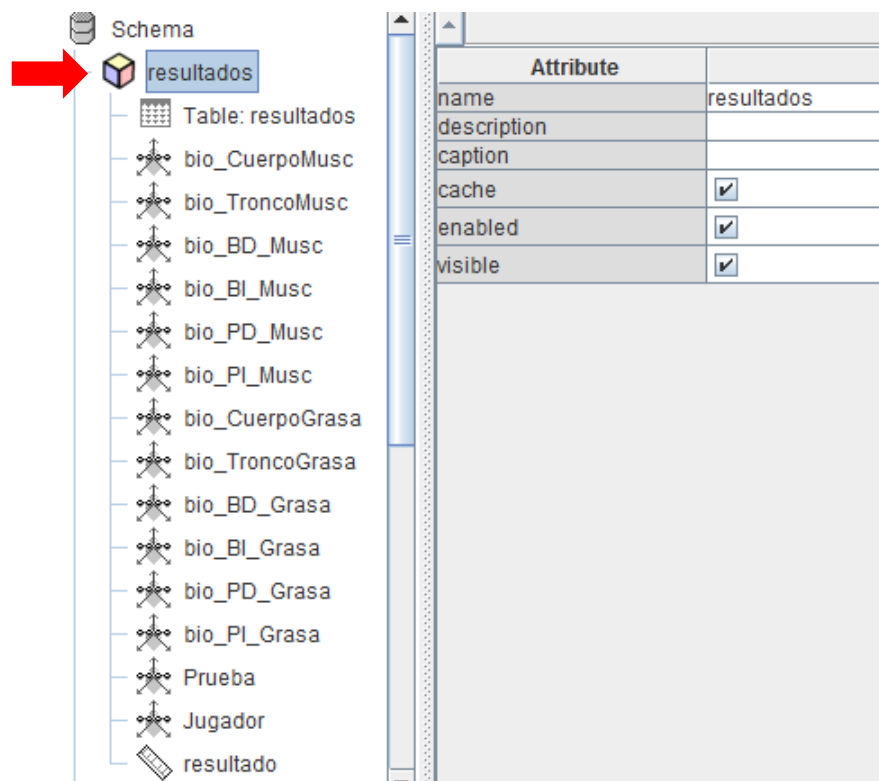
**Figura 14.** Consola de MySQL mostrando la tabla creada.



## 2.3 Definición del cubo

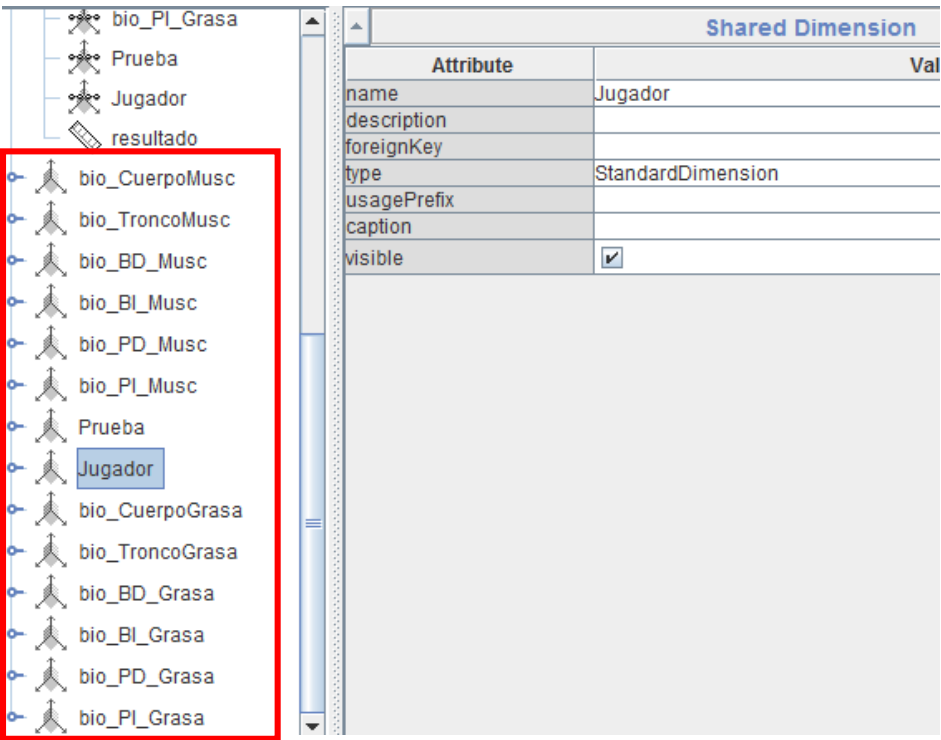
Teniendo disponible la estructura de la base de datos se procede a la creación de la estructura del cubo OLAP. Para ello, Pentaho proporciona la herramienta Pentaho Schema Workbench, es una herramienta gráfica que permite crear y probar esquemas de cubos OLAP. Esta crea un fichero .xml donde se describen las relaciones entre las dimensiones, los niveles de jerarquía de las dimensiones, los hechos y las medidas del cubo, así como la conexión a la base de datos donde se encuentran los datos que van a formar el cubo.

En primer lugar, se definen las conexiones de Pentaho Schema Workbench con la base de datos para la obtención de sus campos. A continuación, se crea un Schema en el cual es añadido el cubo (**Figura 15**).



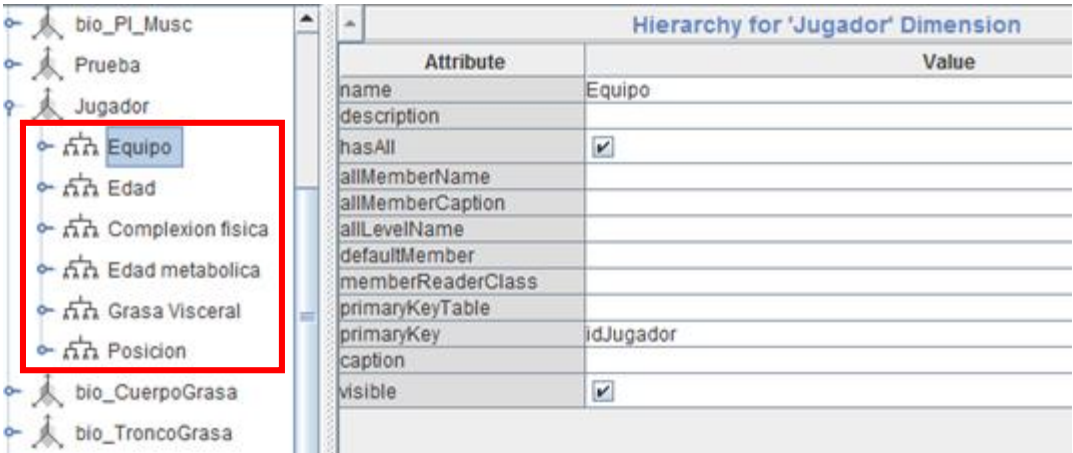
**Figura 15.** Cubo creado dentro del Schema.

Seguidamente son añadidas cada una de las dimensiones del cubo (**Figura 16**).



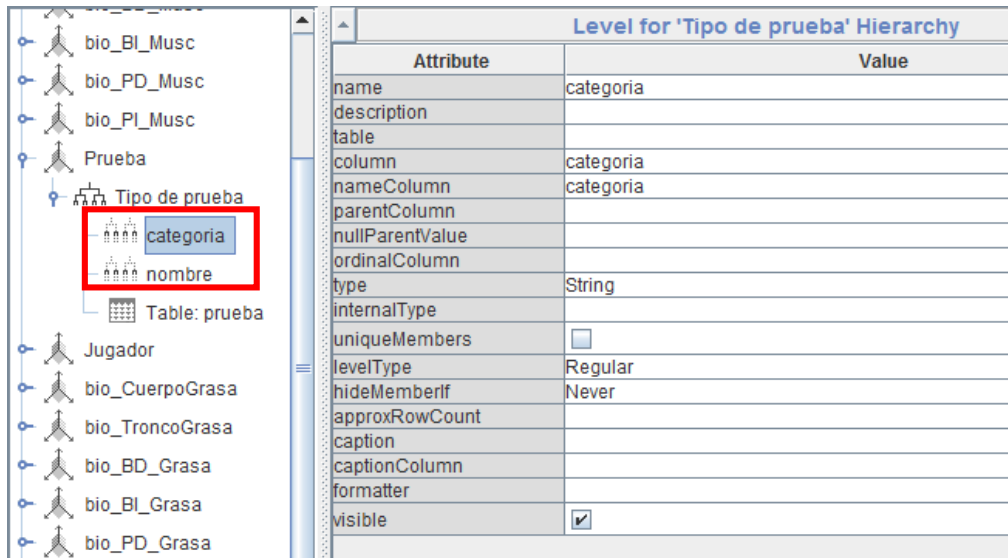
**Figura 16.** Dimensiones del cubo.

Cada dimensión con sus jerarquías. Las jerarquías son niveles de análisis y detalles de la información del modelo multidimensional. Estas jerarquías permitirán más adelante realizar el análisis y la navegación por los datos. Se pueden tener tantas jerarquías como sea necesario dentro de la dimensión, a la hora del análisis se seleccionará la que se desee (**Figura 17**).



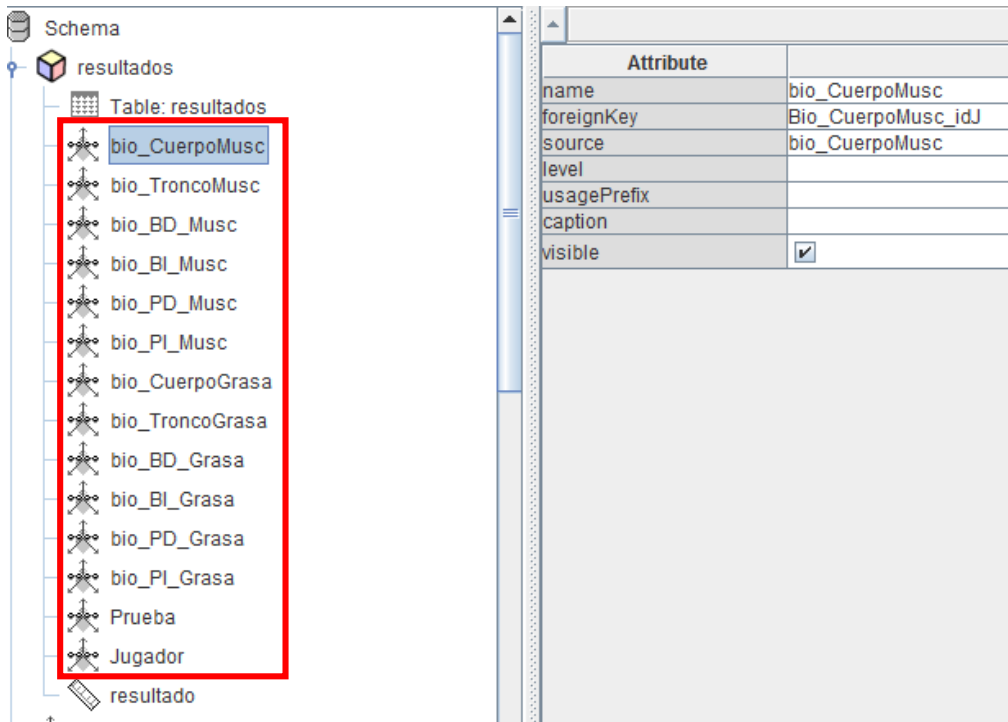
**Figura 74.** Jerarquías de la dimensión Jugador.

Y sus correspondientes niveles de la jerarquía. El orden con el que se crean determina la estructura de la jerarquía (**Figura 18**).



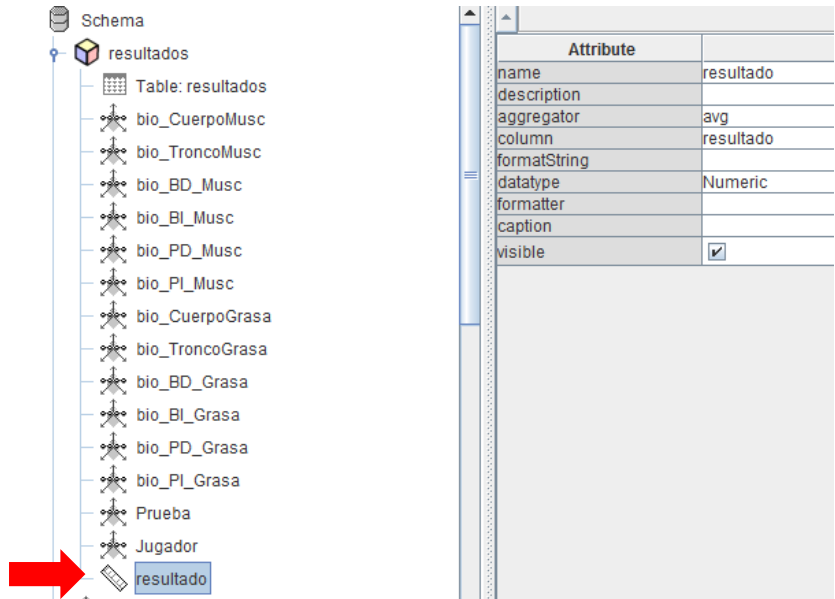
**Figura 18.** Niveles de la jerarquía Tipo de prueba de la dimensión Prueba.

Una vez se tienen todas las dimensiones, son creados los usos de las dimensiones que se relacionan con la tabla de hechos. El cubo hereda todas las características que hayan sido incluidas en las dimensiones previamente creadas, incluyendo todas las jerarquías y sus correspondientes atributos (**Figura 19**).



**Figura 19.** Usos de las dimensiones dentro del cubo.

Como último paso de la creación del cubo se definen las medidas, estas serán los valores de análisis, en este caso como medida se ha seleccionado el campo resultado (**Figura 20**).



**Figura 20.** Medida del cubo.

Como se ha nombrado anteriormente, este trabajo es guardado en un fichero .xml en el cual están representados los esquemas dimensionales, que como se verá más adelante se carga en Pentaho como estructura del cubo. En esta estructura serán representados los datos del almacén.

## 2.4 Proceso ETL

Para llevar a cabo la población de la base de datos se han realizado los procesos de extracción, transformación y carga de los datos. Se disponía de un documento Excel con todos los datos de los jugadores el cual se ha tenido que procesar para cargar dichos datos en la base de datos. Para la realización de estas tareas Pentaho ofrece la herramienta Spoon, también conocido como Kettle. Es una interfaz gráfica de usuario que permite la creación de dichos procesos.

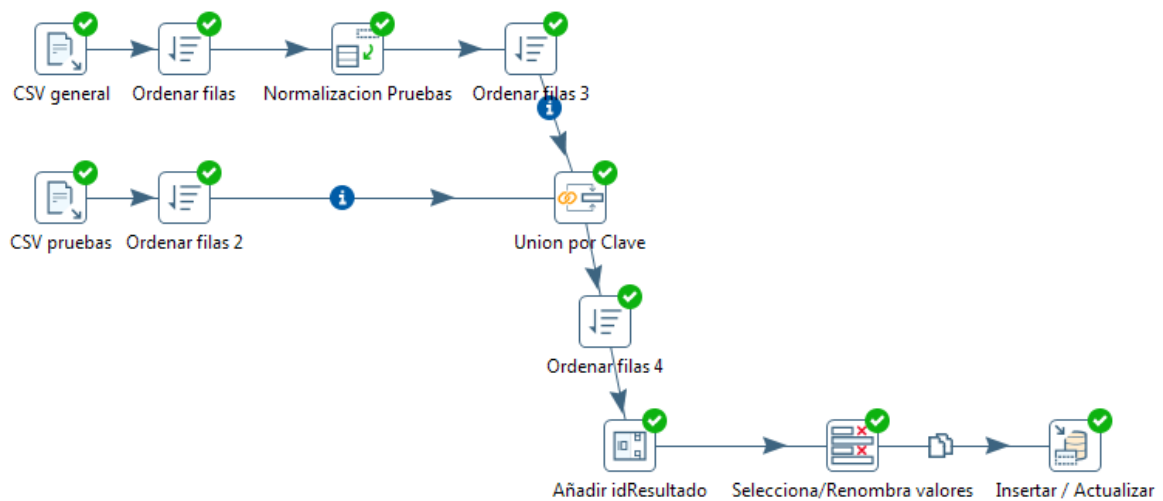
Como en la mayoría de herramientas que ofrece Pentaho, lo primero que se ha de hacer es conectarla con la base de datos donde se quieren guardar los datos transformados.

Ha sido realizada una transformación por cada una de las tablas de la base de datos. Para la mayoría de las tablas se han seguido los pasos que se encuentran en la **Figura 21**. Como en todas las transformaciones en primer lugar se ha importado el CSV donde se encuentran los datos, una vez obtenidos los datos del CSV, se han seleccionado los que requiere esa específica tabla y renombrado para que coincida el nombre de las variables con el nombre dado en la base de datos. Para por último se insertan o actualizan los datos en la tabla seleccionada de la base de datos con la que se ha realizado la conexión.



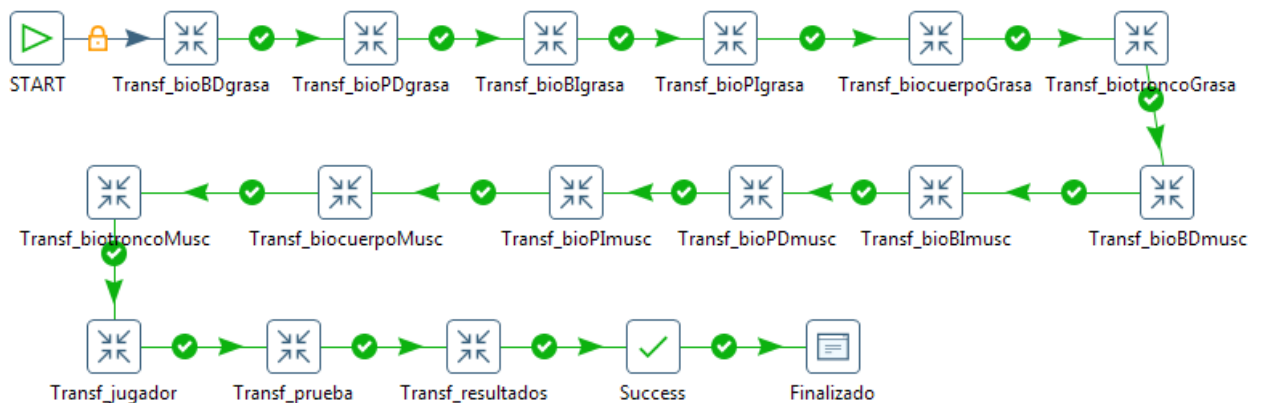
**Figura 21.** Transformación de la tabla.

La transformación más costosa ha sido la de la tabla de hechos, (**Figura 22**), puesto que combina datos de dos CSVs, y, además, se ha tenido que normalizar uno de ellos ya que cada jugador debe tener una fila por cada una de las pruebas realizadas para poder llevar a cabo el análisis.



**Figura 22.** Transformación de la tabla.

Una vez se tienen todas las trasformaciones se genera un trabajo (**Figura 23**), que reúne las transformaciones creadas y las ejecuta de manera ordenada respetando las claves de las distintas tablas.



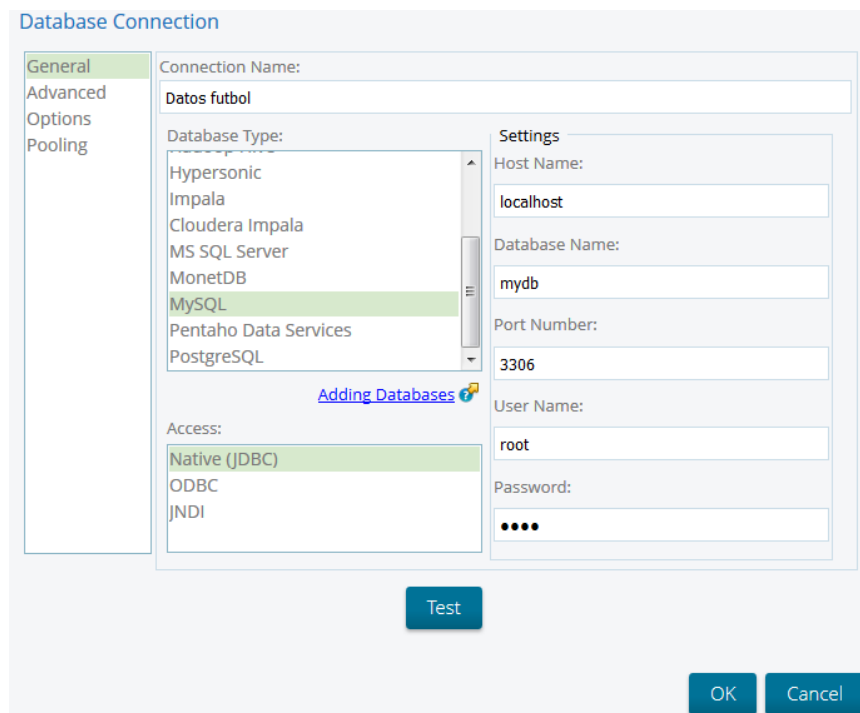
**Figura 23.** Estructura Trabajo.

Finalmente, cuando el trabajo se ha ejecutado con éxito se podrá comprobar que la base de datos ha sido poblada con los datos indicados.

## 2.5 Publicación del cubo en Pentaho

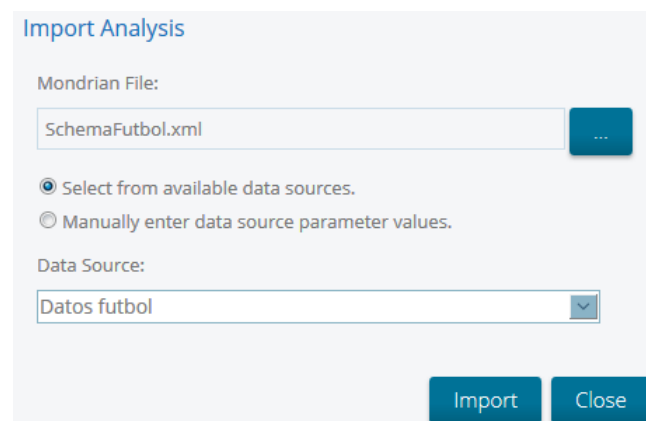
Una vez los pasos anteriores han sido realizados con éxito, es decir, la base de datos haya sido creada y poblada y se disponga de un esquema de la estructura del cubo, se procede a iniciar el servidor de Pentaho para la publicación del cubo en él. Escribiendo la siguiente ruta en el navegador “localhost:8080/pentaho” se accede a dicho servidor.

Como en las demás aplicaciones se ha de crear una conexión con la base de datos, para ello es añadida una conexión JDBC (**Figura 24**).



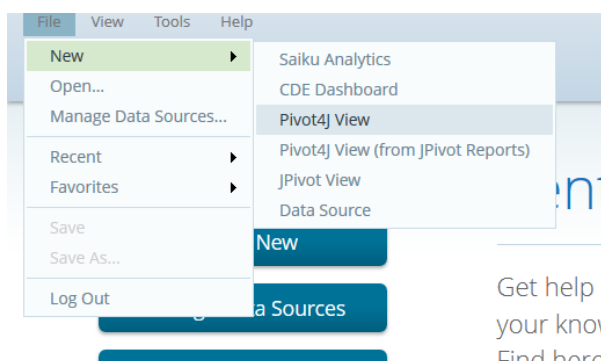
**Figura 24.** Conexión con la BD.

También es importado el esquema del cubo y conectado con la base de datos JDBC para poder ver las tablas representadas (**Figura 25**).



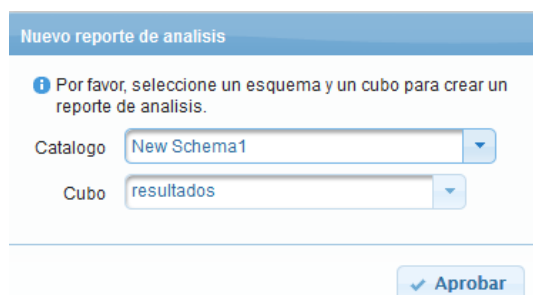
**Figura 25.** Creación de vista análisis.

Una vez ha sido publicado, se procede a la creación de una vista “Pivot4J” (**Figura 26**).



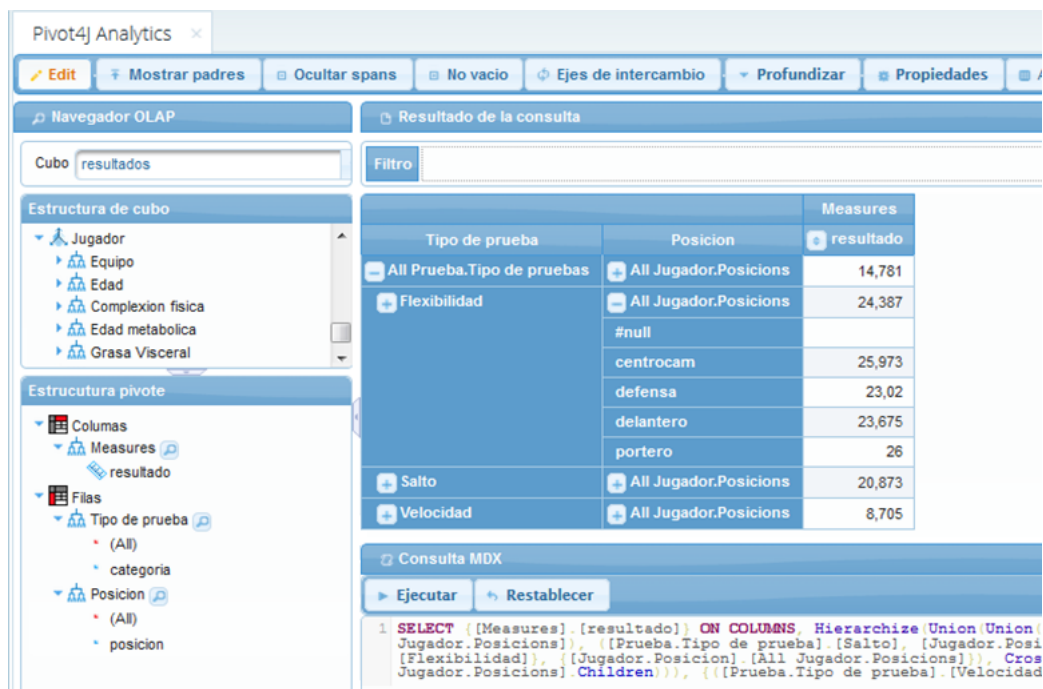
**Figura 26.** Creación de vista Pivot4J.

Se selecciona el cubo del cual se quiere realizar el análisis (**Figura 27**).



**Figura 27.** Selección del cubo.

Y finalmente se genera una visa como la de la **Figura 28**. Esto es lo que el cliente finalmente va a disponer para realización del análisis, pudiendo personalizar el cubo arrastrando los parámetros a las columnas o filas en función a sus consultas.



**Figura 28.** Vista del cubo.



## 2.6 Consultas

Una vez publicado el cubo en Pentaho, el producto está listo para ser entregado al cliente, permitiéndole realizar las consultas que desee con un panel como el de la **Figura 29**.

The screenshot shows the Pivot4J Analytics interface. The top bar includes buttons for 'Edit', 'Mostrar padres', 'Ocultar spans', 'No vacío', 'Ejes de intercambio', 'Profundizar', and 'Propiedades'. Below this is the 'Navegador OLAP' section with a 'Cubo' dropdown set to 'resultados' and a 'Filtro' field. The main area is divided into three zones:

- Zona 1 (Estructura de cubo):** A tree view showing dimensions: 'Jugador' (with sub-items 'Equipo', 'Edad', 'Complexion fisica', 'Edad metabolica', 'Grasa Visceral') and 'Medidas'.
- Zona 2 (Estructura pivote):** A tree view showing the pivot structure: 'Columnas' (with 'Medidas' and 'resultado') and 'Filas' (with 'Tipo de prueba' and 'Posicion').
- Zona 3 (Resultado de la consulta):** A table displaying the query results. The table has columns for 'Tipo de prueba', 'Posicion', and 'Medidas' (with a sub-column 'resultado').

The query results table is as follows:

Tipo de prueba	Posicion	Medidas (resultado)
All Prueba.Tipo de pruebas	All Jugador.Posicions	14,781
Flexibilidad	All Jugador.Posicions	24,387
	#null	
	centrocam	25,973
	defensa	23,02
	delantero	23,675
	portero	26
Salto	All Jugador.Posicions	20,873
Velocidad	All Jugador.Posicions	8,705

At the bottom, there is a 'Consulta MDX' section with an 'Ejecutar' button and a 'Restablecer' button. Below these buttons is the MDX query text:

```
1 SELECT {[Measures].[resultado]} ON COLUMNS, Hierarchize(Union(Union(Jugador.Posicions), {[Prueba.Tipo de prueba].[Salto], [Jugador.Posi[Flexibilidad], {[Jugador.Posicion].[All Jugador.Posicions]}}, CrosJugador.Posicions] Children)), {[Prueba.Tipo de prueba].[Velocidad]
```

**Figura 29.** Ejemplo de una vista del cubo.

Para la realización de consultas se deben manejar 3 zonas. En la estructura del cubo (Zona 1), se dispone de las medidas y dimensiones que se han creado para llevar a cabo su análisis. Las necesarias que se consideren necesaria para la consulta se han de arrastrar hasta las columnas o filas de la zona de estructura del pivote (Zona 2). Dependiendo de cuales sean seleccionadas y como sean ordenadas se creará un resultado u otro.

En la zona 3 es donde será mostrado el resultado de la consulta pudiendo ampliar o agrupar las filas y columnas en función de lo que se desee visualizar.

### 3. Análisis

Para proceder al análisis de los datos se ha hecho uso de 3 herramientas: el cubo de análisis que ofrece Pentaho, el complemento de Excel Power Pivot para una visualización gráfica y una herramienta de minería de datos llamada Weka.

#### 3.1 Pentaho

El cubo publicado en Pentaho es una gran herramienta de análisis, está pensado para ser utilizado por analistas para validar y descartar hipótesis de forma rápida, no es necesario que sean profesionales de la información por lo que es muy fácil de utilizar. Su estructura es completamente editable, pudiendo elegir las características a analizar simplemente arrastrándolas hacia sus ejes.

Se han realizado diferentes configuraciones del cubo para su análisis. Primero se han representado para cada prueba la media de los resultados obtenidos para los distintos equipos. Como se había escrito con anterioridad los equipos están divididos por las distintas edades, su clasificación de mayor a menor edad es: 1er equipo, cadete A, cadete B, infantil C, alevín A, alevín B, alevín C, benjamín A, benjamín B, benjamín C. En la **Figura 30** puede observarse la diferencia de resultados entre los equipos formado por jugadores jóvenes y los equipos formados por jugadores más adultos. Los cadetes alcanzan el mayor resultado en la prueba de Salto CMJ, más de 33 centímetros, mientras que los alevines obtienen el menor resultado.

Salto CMJ	All Jugador.Equipos	24,747
	1er equipo	
	alevin A	17,413
	alevin B	20,952
	alevin C	20,035
	benjamin A	17,567
	benjamin B	20,191
	benjamin C	18,333
	cadete A	33,25
	cadete B	30,077
	infantil C	24,925

**Figura 30.** Consulta del cubo, resultados del salto CMJ por equipos.

La **Figura 31** muestra el tiempo medio que han tardado los jugadores de cada equipo en realizar 6 series de 20 metros. Al igual que para las pruebas de salto, los jugadores pertenecientes a equipos con mayor edad obtienen mejores resultados. El equipo que ha tardado menor tiempo medio en realizar la prueba ha sido el primer equipo, mientras que los que más tiempo han tardado han sido los benjamines, equipo compuesto por los jugadores de menor edad.

TIEMPO 6X20 MEDIA	All Jugador.Equipos	3,511
	1er equipo	2,855
	alevin A	3,67
	alevin B	3,947
	alevin C	3,908
	benjamin A	4,211
	benjamin B	4,114
	benjamin C	4,01
	cadete A	2,971
	cadete B	3,078
	infantil C	3,418

**Figura 31.** Consulta del cubo, resultados del tiempo medio en 6x20 por equipos.

En la **Figura 32**, se representa el porcentaje de velocidad perdida en las series de la prueba anterior. Los equipos benjamines a la vez que han sido el equipo que mayor tiempo ha tardado en realizarla también son los que mayor porcentaje de velocidad han perdido, son jugadores que aún no tienen la capacidad de medir su resistencia para repartirla a lo largo de la prueba. Por el contrario, sujetos del equipo benjamín C han sido los que menos velocidad han perdido, esto puede ser debido a que, al no tener una gran velocidad, tampoco tienen mucho que perder.

% PERDIDA VELOCIDAD 6X20	All Jugador.Equipos	8,639
	1er equipo	6,138
	alevin A	8,77
	alevin B	9,209
	alevin C	7,32
	benjamin A	11,148
	benjamin B	14,981
	benjamin C	3,04
	cadete A	5,488
	cadete B	6,961
	infantil C	9,692

**Figura 32.** Consulta del cubo, resultados de % de velocidad perdida por equipos.

Por otra parte, se han representado las pruebas agrupadas por las distintas posiciones de los jugadores, para comprobar si existen grandes diferencias entre ellas. La **Figura 33** y la **Figura 34** representan la flexibilidad y la agilidad respectivamente. Los porteros obtienen los mejores resultados en ambas pruebas.

Flexibilidad	All Jugador.Posicions	24,387
	centrocam	25,973
	defensa	23,02
	delantero	23,675
	portero	26

**Figura 33.** Consulta del cubo, resultados de flexibilidad por posiciones.

Agilidad	All Jugador.Posicions	15,618
	#null	
	centrocam	15,495
	defensa	15,737
	delantero	15,296
	portero	16,741

**Figura 34.** Consulta del cubo, resultados de agilidad por posiciones.

A continuación, la **Figura 35** representa los distintos saltos agrupados por posiciones. Se puede observar como en todas las pruebas el centrocampista es el que más centímetros salta.

Salto CMJ	All Jugador.Posicions	24,747
	centrocam	26,501
	defensa	24,209
	delantero	23,793
	portero	22,652
Salto CMJ + BRAZOS	All Jugador.Posicions	28,608
	centrocam	30,425
	defensa	27,741
	delantero	27,771
	portero	27,396
Salto SJ	All Jugador.Posicions	22,383
	centrocam	23,85
	defensa	21,53
	delantero	22,045
	portero	20,891

**Figura 35.** Consulta del cubo, resultados de pruebas de salto agrupados por posiciones.

En la **Figura 36** se representa el porcentaje de elasticidad y la coordinación, la elasticidad debe de estar por encima de 10% para considerarse un buen resultado y la coordinación por encima de 12. Los defensas obtienen el mayor porcentaje de elasticidad, mientras que los porteros son la posición con mayor coordinación. Tanto los delanteros como los porteros deben mejorar su porcentaje de elasticidad ya que no se encuentran dentro de unos valores correctos.

% ELAST.	All Jugador.Posicions	11,851
	#null	
	centrocam	11,883
	defensa	14,637
	delantero	9,539
	portero	6,556
	All Jugador.Posicions	16,778
	#null	
	centrocam	16,849
	defensa	14,45
	delantero	17,585
	portero	25,01

**Figura 36.** Consulta del cubo, resultados de pruebas de salto agrupados por posiciones.

Por último, la **Figura 37** representa las distintas pruebas de velocidad. El centrocampista es el tipo de jugador que en menos tiempo las realiza, por lo tanto, son los jugadores más rápidos. Por otra parte, los defensas son los más lentos en la prueba de 10 metros, mientras que los porteros son los más lentos en las pruebas de 20 metros y de 6x20 metros.

TIEMPO10M	All Jugador.Posicions	1,7
	#null	1,631
	centrocam	1,695
	defensa	1,734
	delantero	1,713
	portero	1,712
TIEMPO 20M	All Jugador.Posicions	3,399
	#null	2,875
	centrocam	3,443
	defensa	3,579
	delantero	3,448
	portero	3,684
TIEMPO 6X20 MEDIA	All Jugador.Posicions	3,511
	#null	2,855
	centrocam	3,538
	defensa	3,615
	delantero	3,604
	portero	3,868

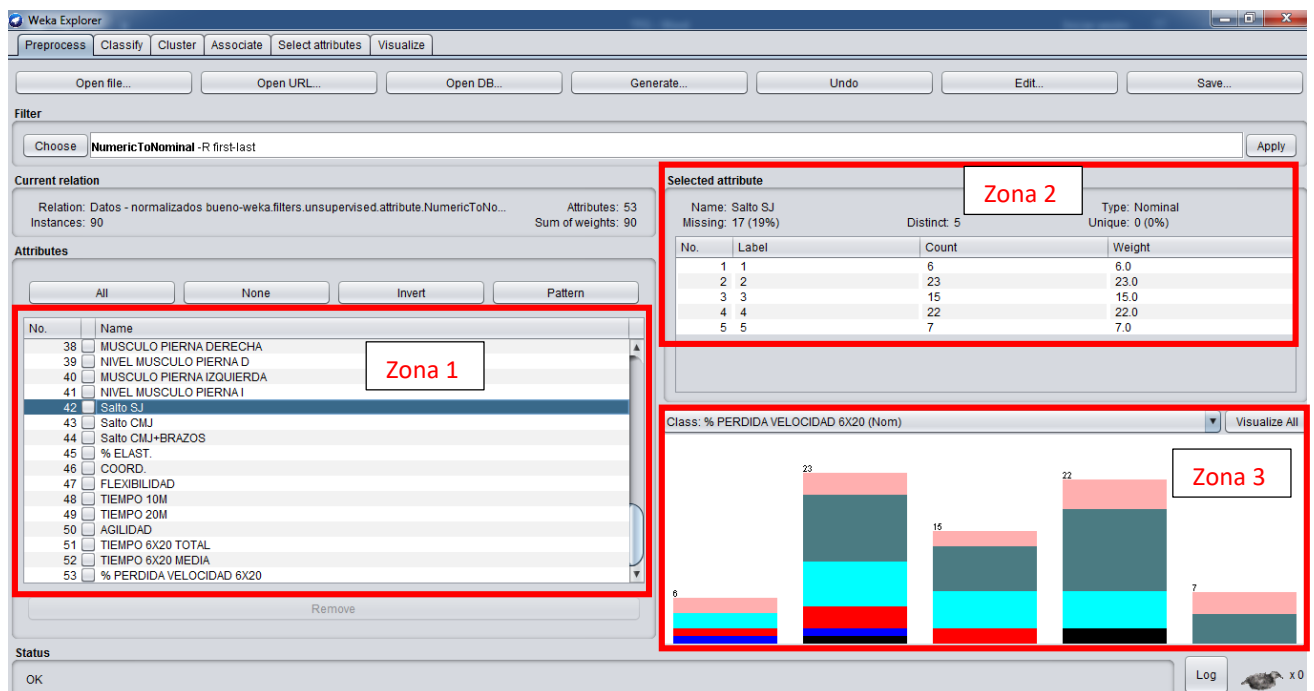
**Figura 37.** Consulta del cubo, resultados de pruebas de velocidad agrupado por posiciones.

Las características que definen cada posición son, coordinación y flexibilidad en los porteros, la potencia de salto y velocidad para los centrocampistas y la elasticidad en los defensas.

Como hemos visto, el cubo es una herramienta dinámica y sencilla, una vez el usuario dispone del cubo, si surgiera la necesidad de la realización de un nuevo análisis con otros parámetros, él mismo podría hacerlo, ya que una vez el cubo ha sido publicado es muy sencillo realizar nuevos análisis, no se necesita ser profesional de las tecnologías de la información para su utilización, simplemente hay que tener claro que se quiere analizar y plasmarlo en los ejes del cubo. El objetivo de esta herramienta es que cualquier analista pueda utilizarla y realizar un gran análisis de una forma sencilla y eficaz.

## 3.2 Weka

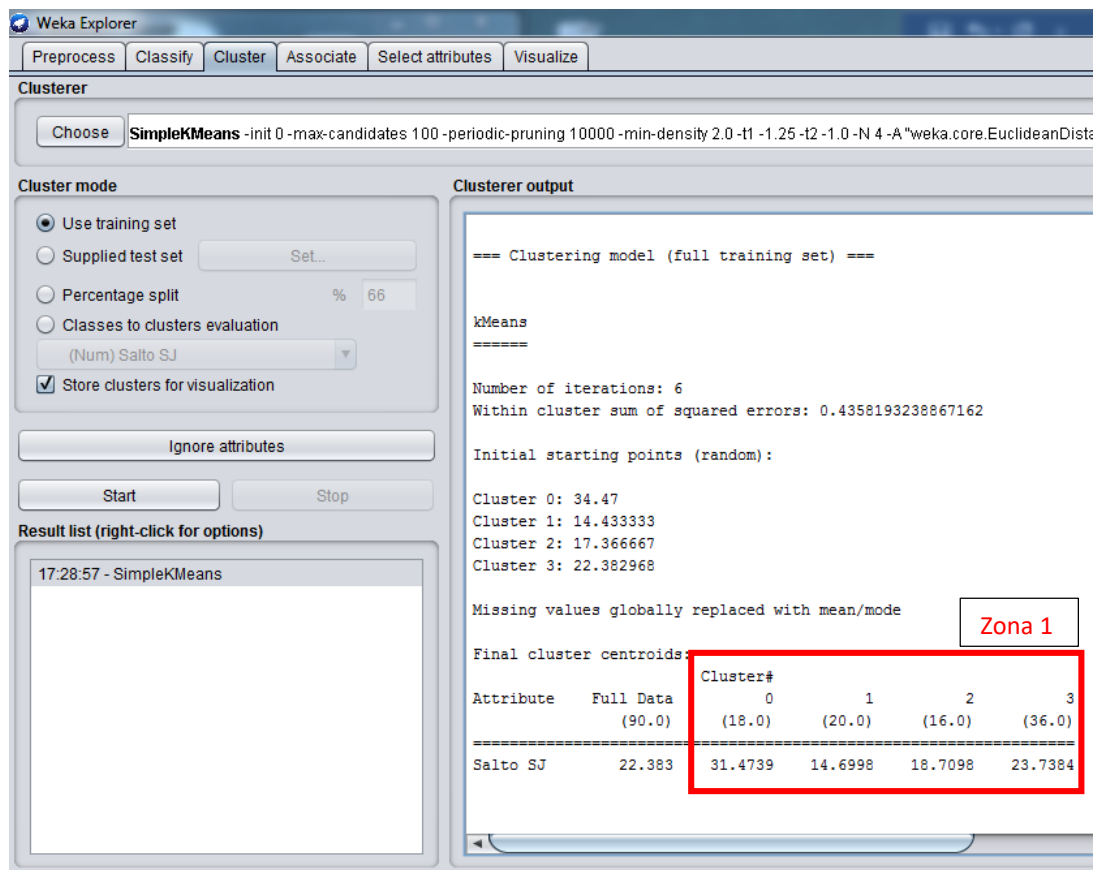
Weka es una herramienta con un conjunto de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos se pueden aplicar directamente sobre un conjunto de datos o bien pueden ser llamados desde código Java. Contiene herramientas para procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. Todas ellas unidas bajo una interfaz gráfica que facilita mucho su uso (**Figura 38**).



**Figura 38.** Ejemplo de la pantalla principal de Weka.

La **Figura 38** muestra la pantalla principal de Weka. Primero es cargado un CSV con los datos a analizar y son mostrados en la Zona 1, donde pueden ser eliminados los innecesarios, así como puede ser aplicado algún tipo de filtro. Seleccionando uno de los atributos se observa en la Zona 2 los detalles de los datos del atributo seleccionado y en la zona 3 muestra una representación gráfica de los datos de dicho atributo.

Se han utilizado varios algoritmos de Weka. Primero, para hacer uso de un algoritmo de árbol de decisión, se han tenido que discretizar los datos. Para esto, se ha utilizado el algoritmo K-Medias (**Figura 39**), su función es delimitar unos valores en los que se agruparan los datos a la hora de discretizarlos.



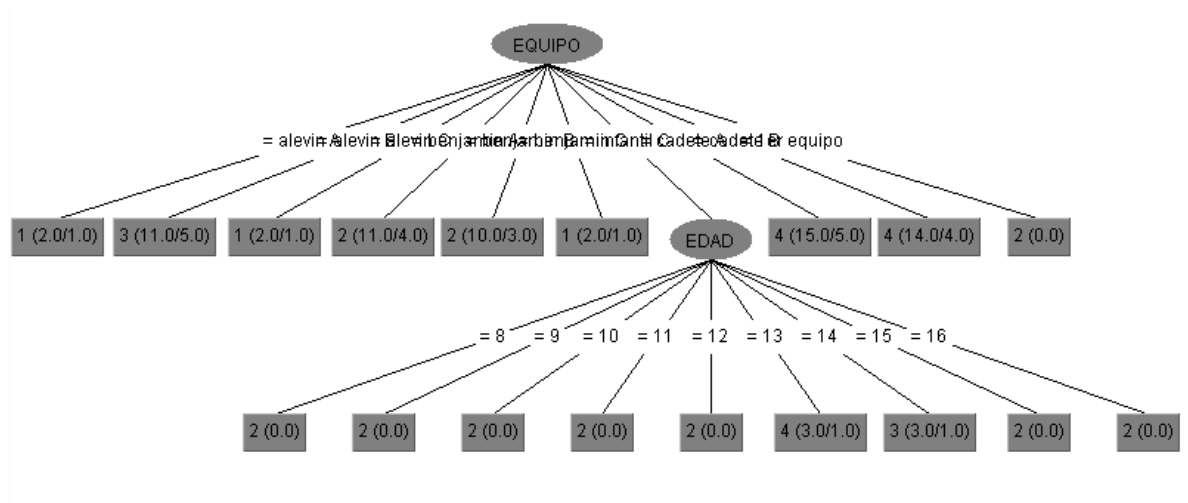
**Figura 39.** Clusters obtenidos por algoritmo K-Medias para el atributo Salto SJ.

Para este ejemplo del atributo Salto SJ, en la Zona 1 se observan los valores que el algoritmo K-Medias ha obtenido para la delimitación de los grupos.

Una vez se dispone de los clústeres en los que agrupar, se han discretizado los datos y se han cargado en Weka los nuevos datos para proceder de nuevo a su análisis.

Se ha ejecutado el algoritmo J48, para cada uno de los atributos referentes a las pruebas físicas realizadas a los jugadores. El algoritmo J48 es un árbol de decisión el cual indica cuales son los atributos que ha utilizado para clasificar los datos de una variable indicada. En la **Figura 40** se observa el árbol de decisión resultante de aplicar el algoritmo J48 sobre la prueba de Salto SJ.





**Figura 40.** Árbol resultante de la aplicación del algoritmo J48 al atributo Salto SJ.

Los atributos que ha utilizado para su clasificación han sido el equipo al que pertenece el jugador y su edad. La elección de estas dos variables indica que a medida que el jugador crece la fuerza física necesaria para la realización del Salto SJ aumenta, ya que los equipos están divididos por edades.

Se ha ejecutado el algoritmo J48 sobre las distintas pruebas, los clasificadores que han mostrado los árboles resultantes se han descrito en la **Tabla 3**.

**Tabla 3.** Clasificadores seleccionados por el algoritmo J48 para cada prueba.

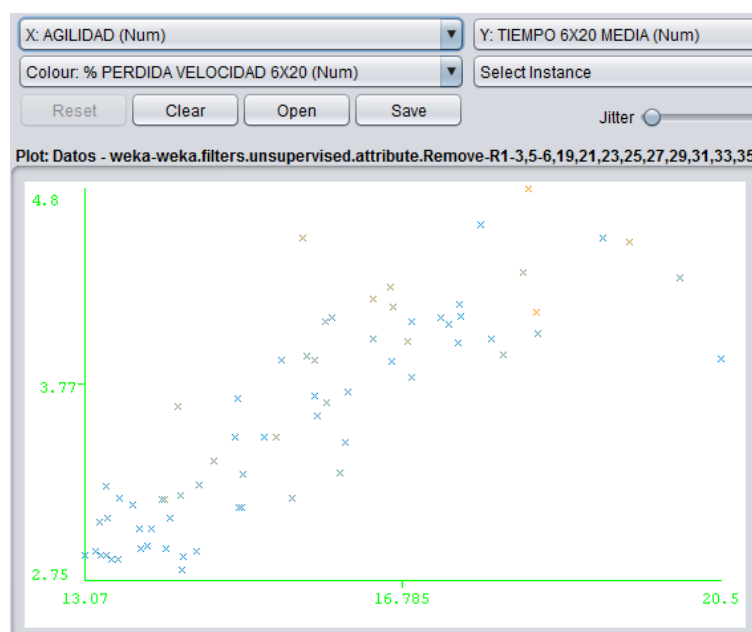
PRUEBA	CLASIFICADORES
Salto SJ	Equipo – Edad
Salto CMJ	Peso
Salto CMJ + brazos	Peso
% Elasticidad	Peso
Coordinación	
Flexibilidad	
Tiempo en 10 metros	Equipo – Masa ósea – Complexión física – Posición
Tiempo en 20 metros	Equipo
Agilidad	Equipo – N° calzado – Edad
Tiempo en 6x20 total	Equipo – Posición – N° calzado
Tiempo en 6x20 media	Equipo – Posición – N° calzado
% perdida velocidad 6x20	

El peso ha sido determinante para clasificar los datos de las pruebas de Salto CMJ, Salto CMJ + Brazos y el % de elasticidad. Esto puede ser interpretado de manera que, a más peso, más musculatura, por lo que tiene más capacidad de salto, del peso depende el porcentaje de elasticidad.

Para las pruebas de velocidad en 10 metros, 6x20 total y 6x20 media, la posición de los jugadores ha intervenido en su clasificación, lo que refleja que la velocidad de los jugadores varía dependiendo de la posición que desempeñen.

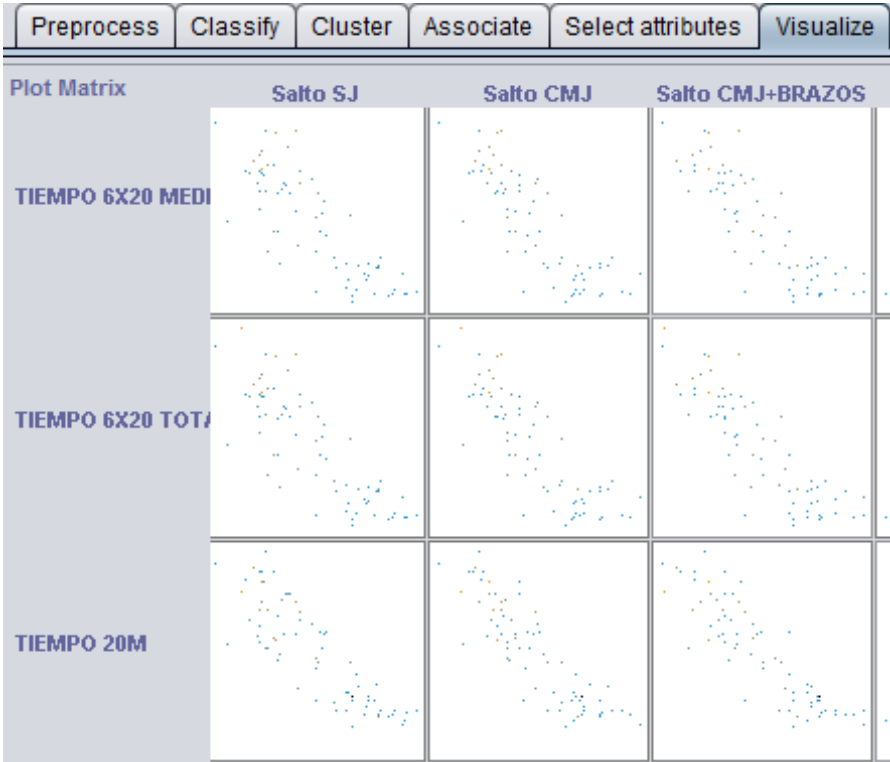
En los resultados existe una gran densidad de los atributos Equipo, Edad y N° Calzado, estos están inmensamente relacionados con la edad del jugador. Debido a esta gran influencia en los resultados sería interesante la realización de evaluaciones centradas en grupos nutridos de jugadores de la misma edad, de manera que estos factores no alteren los resultados y puedan ser observados el resto de factores de la muestra con más claridad.

Por otra parte, Weka también ofrece la oportunidad de ver los atributos representados visualmente en gráficas, esto permite definir si existe correlación entre pares de atributos. Se crean graficas como la de la **Figura 41** donde puede verse representado la agilidad frente a el tiempo medio en hacer 6 series de 20 metros. En esta grafica se observa una correlación entre la agilidad y el tiempo tardado en hacer 6 series de 20 metros, por lo que los jugadores más rápidos también son los más ágiles.



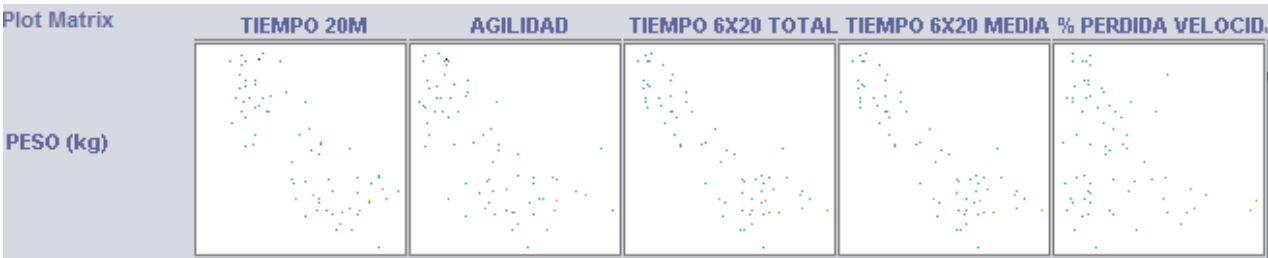
**Figura 41.** Grafica de la agilidad frente al tiempo medio de 6 series de 20 metros.

También permite la visualización en conjunto de varias gráficas, en la **Figura 42** se observa una gran correlación inversa entre las pruebas de salto y las pruebas de velocidad. Lo que indica que los sujetos que más distancia han saltado, menos tiempo han hecho en la prueba de velocidad. Los jugadores más rápidos son a la vez los que saltan más.

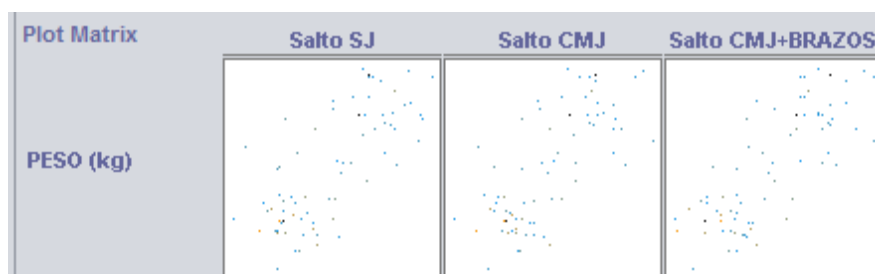


**Figura 42.** Graficas de pruebas de velocidad frente a pruebas de salto.

En las gráficas de la **Figura 43** y la **Figura 44**, se observa que hay una relación entre el peso de los jugadores y los resultados que han obtenido en las pruebas físicas. Los individuos con mayor peso son los que más distancia saltan y mejor resultado obtienen en las pruebas de velocidad y agilidad.



**Figura 43.** Graficas influencia del peso en las pruebas de velocidad y agilidad.

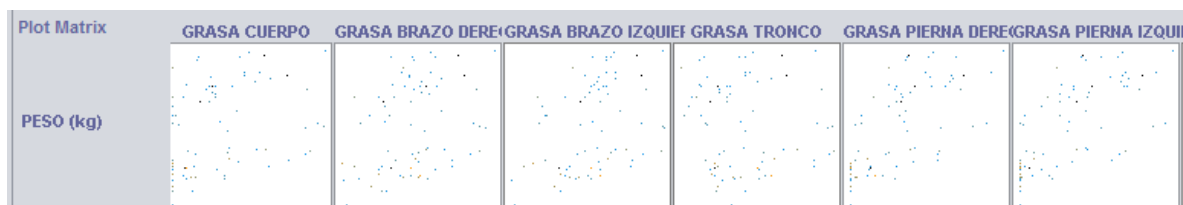


**Figura 44.** Graficas influencia del peso en las pruebas de salto.

Se observa en la **Figura 45** la influencia del musculo en el peso de los jugadores, mientras que como se puede ver en la **Figura 46**, la grasa no tiene influencia. A más peso necesariamente hay más musculo, pero no más grasa, por lo que se deduce que son sujetos con gran masa muscular ya que de su peso depende mucho el musculo.

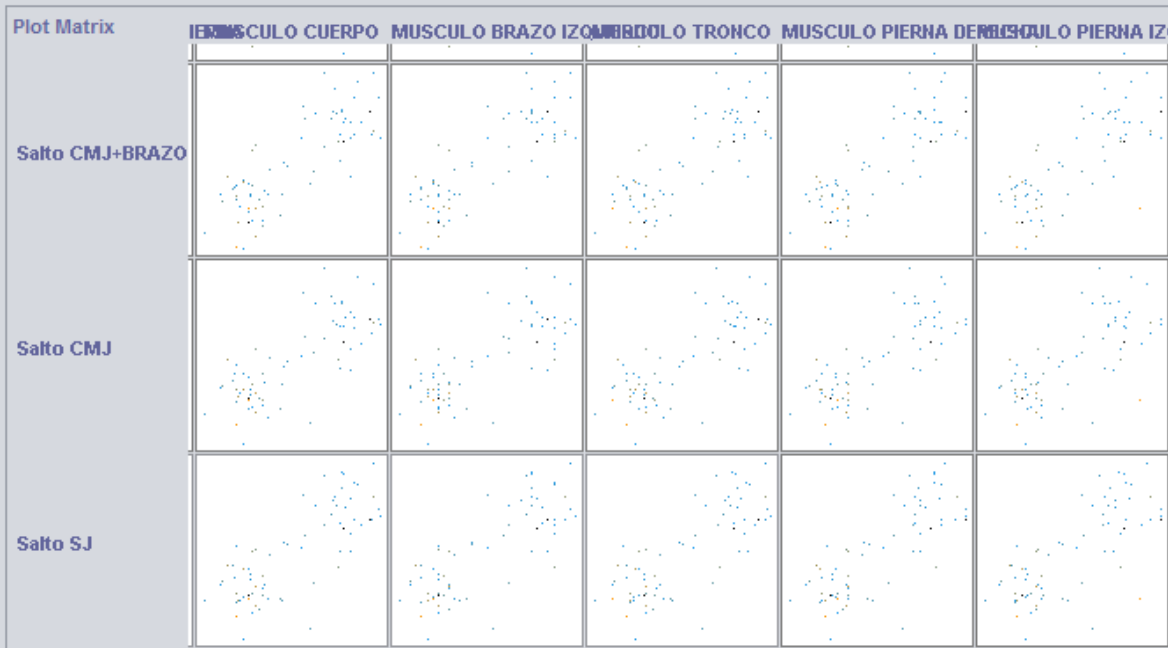


**Figura 45.** Graficas influencia peso en musculo.

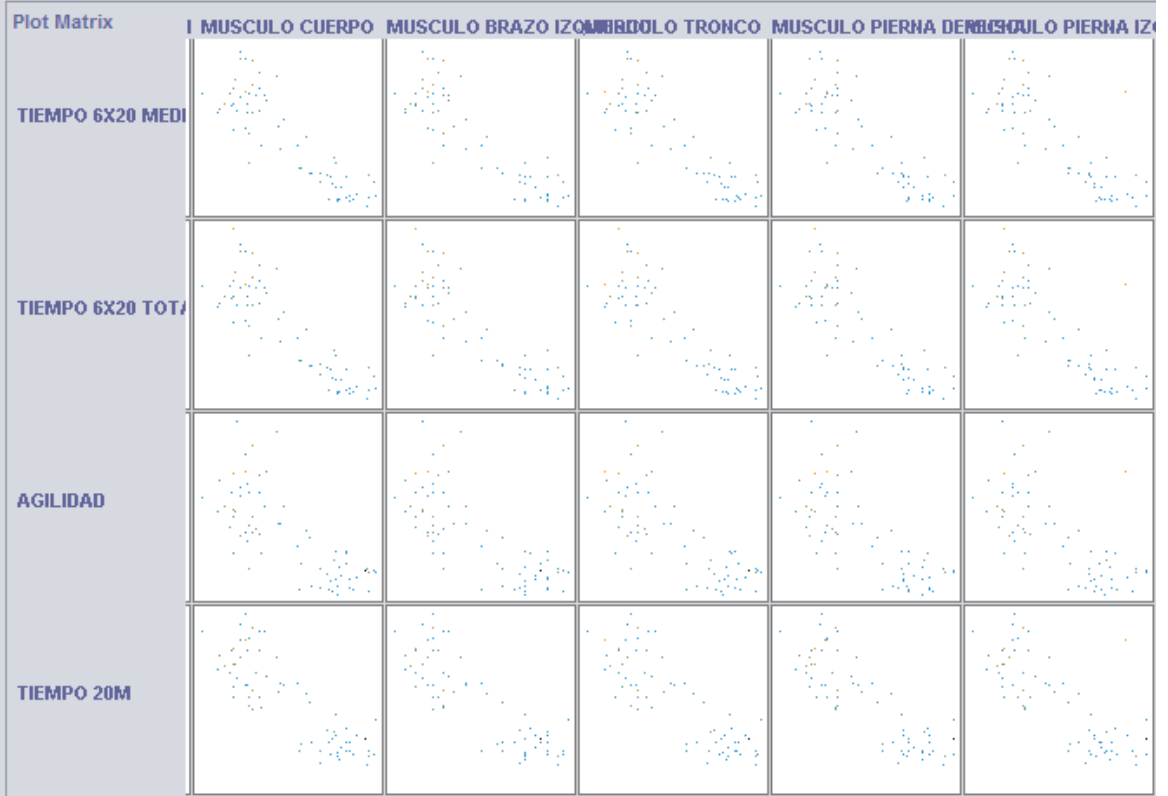


**Figura 46.** Graficas influencia peso en grasa.

A su vez, como se observa en la **Figura 47** y la **Figura 48** existe una correlación entre las distintas pruebas de salto y velocidad y los kilogramos de musculo que tienen los individuos. Cuanta más peso muscular tienen los jugadores, mejores resultados obtienen en las pruebas. Esto también está relacionado con la edad, a más edad, más peso muscular y mejores resultados.

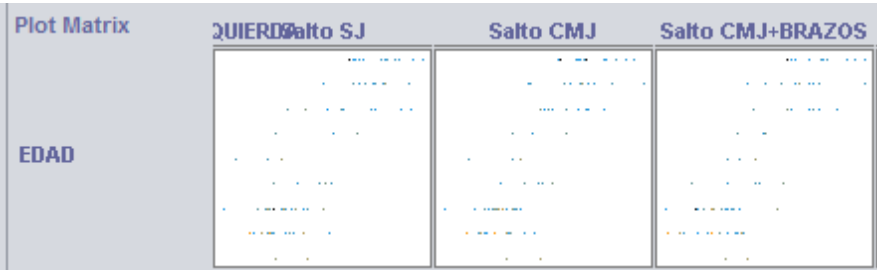


**Figura 47.** Graficas influencia del musculo en pruebas de salto.

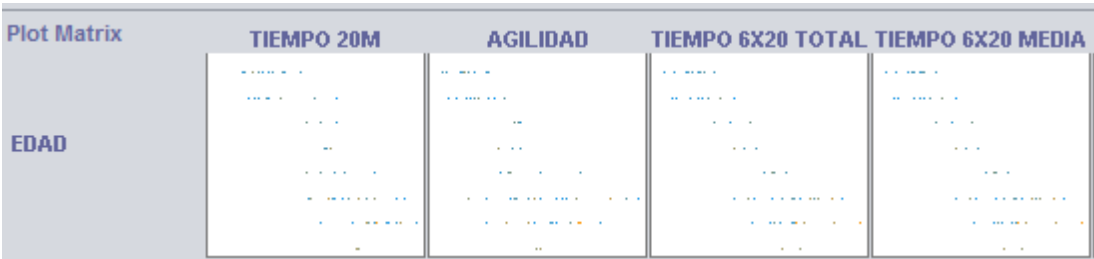


**Figura 48.** Grafica influencia del musculo en las pruebas de velocidad y agilidad.

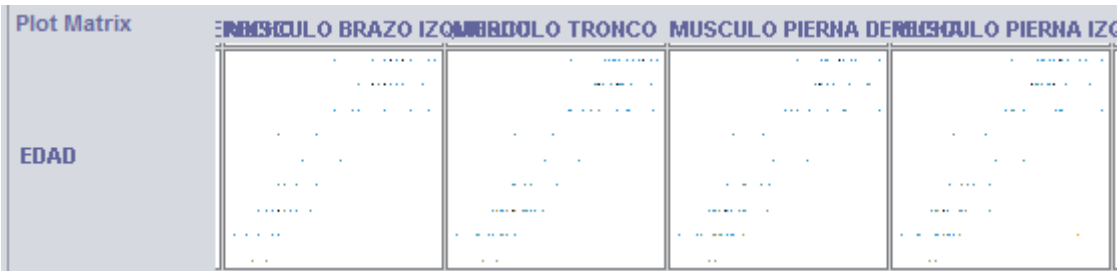
En las **Figuras 49, 50, 51, 52** se observa la influencia de la edad en gran cantidad de parámetros, tanto en las pruebas de salto y velocidad, como en los kilogramos de musculo, así como en la altura, número de calzado y peso. En vista de la gran influencia de la edad de los jugadores, para un futuro análisis sería conveniente realizar evaluaciones centradas en grupos nutridos de jugadores de la misma edad, de manera que haya suficientes datos sin que estos factores alteren los resultados y pueda verse más claramente el efecto del resto de variables en la muestra.



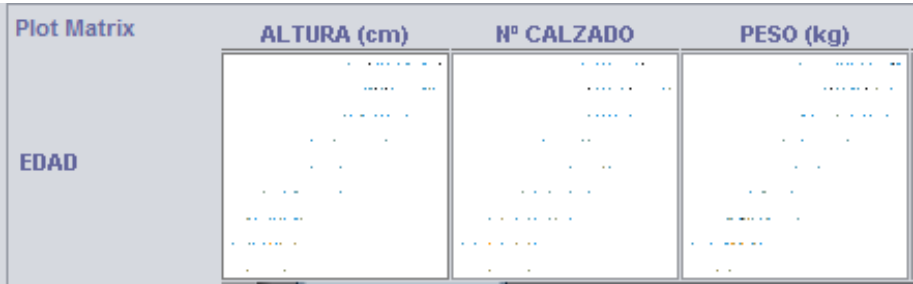
**Figura 49.** Graficas influencia de la edad en pruebas de salto.



**Figura 50.** Graficas influencia de la edad en pruebas de velocidad.



**Figura 51.** Graficas influencia de la edad en el musculo.

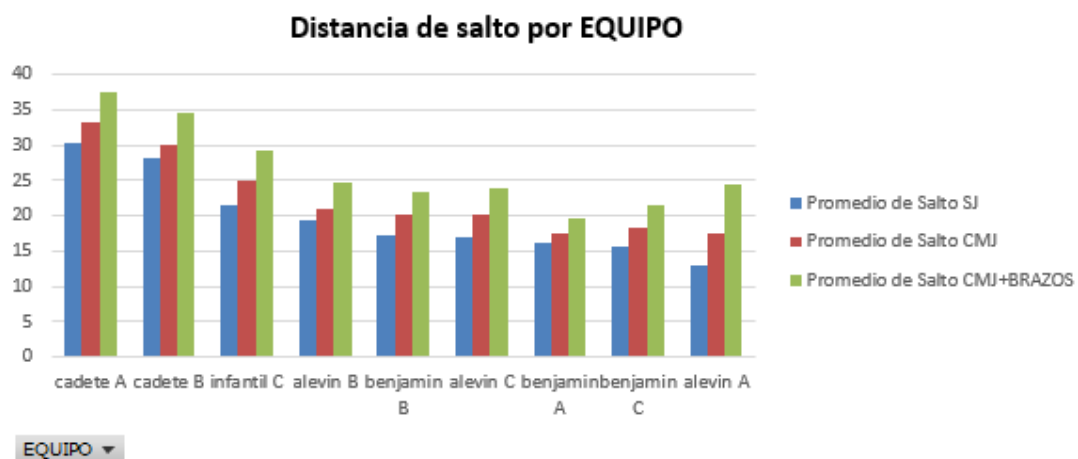


**Figura 52.** Graficas influencia de la edad en la altura, numero de calzado y peso.

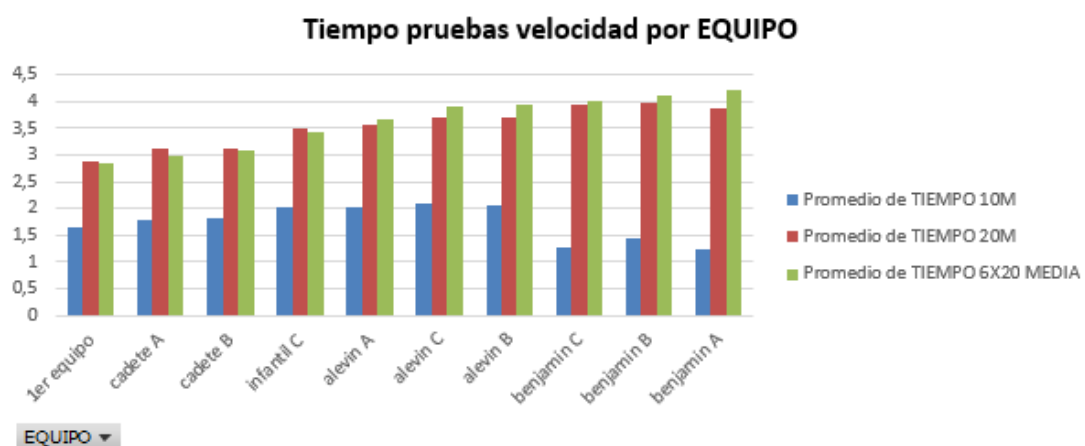
### 3.3 Power Pivot

El complemento de Excel, Power Pivot permite representar la información gráficamente. De esta manera ha sido posible la creación un cuadro de mando el cual con un golpe de vista pueda ser interpretada toda la información. Ha sido seleccionada esta herramienta para la realización del cuadro de mando ya que permite combinar grandes volúmenes de datos, así como realizar análisis de la información rápidamente. Permite la creación de gráficos son dinámicos, al modificar o añadir información, se actualizarán automáticamente, y todos sus atributos pueden ser filtrados o agrupados de manera muy sencilla.

La **Figura 53** y la **Figura 54** representan las pruebas de salto y velocidad respectivamente, estas están agrupadas por los distintos equipos. Se observa en ambas pruebas mejores resultados para los equipos en categorías superiores debido a que los equipos están divididos por edades, por lo que, a mayor edad, mejores resultados se obtienen.

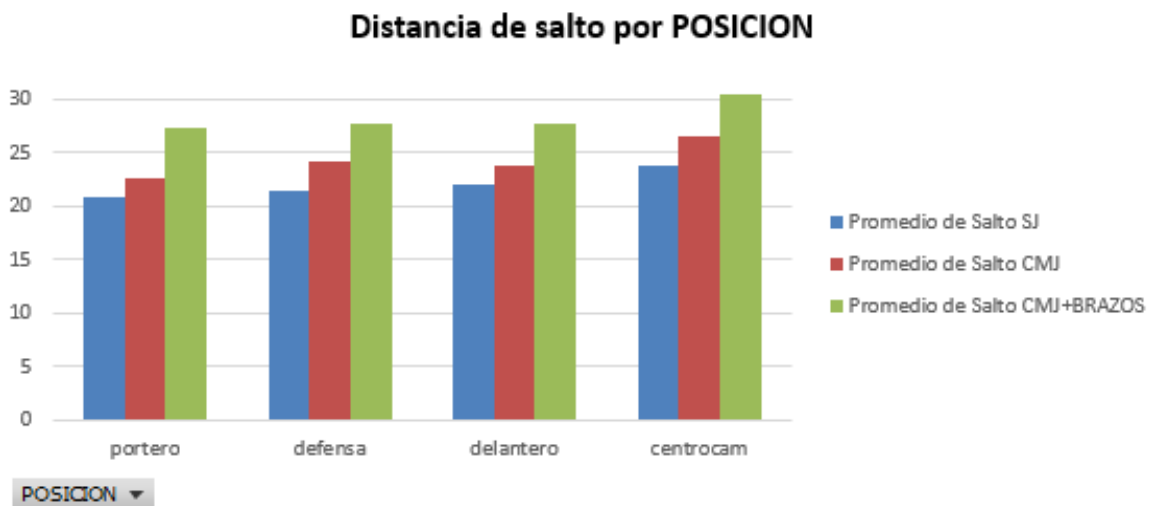


**Figura 53.** Grafica saltos por equipos.

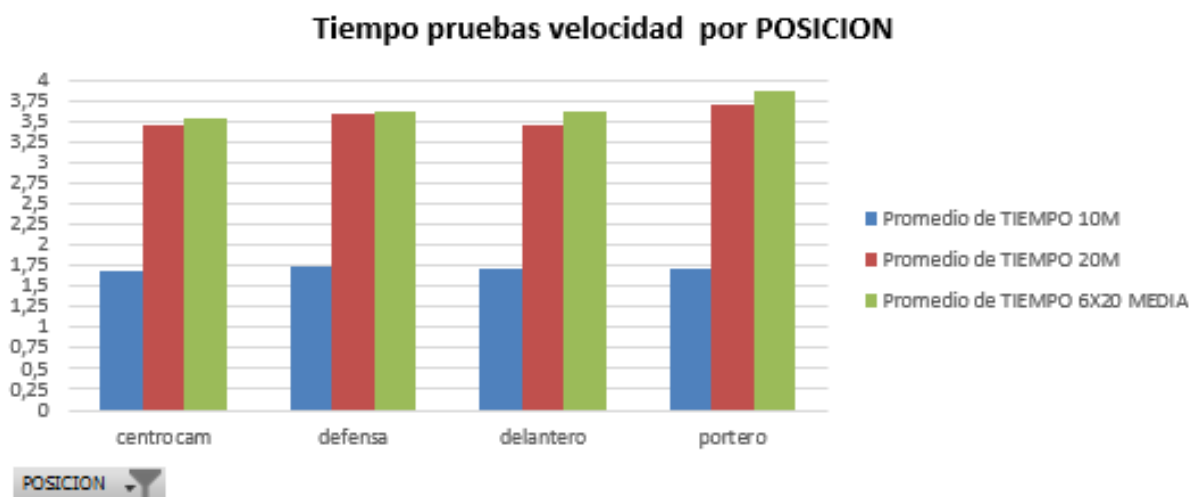


**Figura 54.** Grafica velocidad por equipo.

La **Figura 55** y la **Figura 56** representan las mismas pruebas anteriores, salto y velocidad, pero esta vez agrupado por las distintas posiciones de los jugadores. Tanto en la **Figura 55** como en la **Figura 56** el portero es el que peores resultados obtiene, mientras que el centrocampista es el que mejores resultados obtiene tanto en salto como en velocidad. Los defensas y delanteros obtienen resultados bastante igualados.



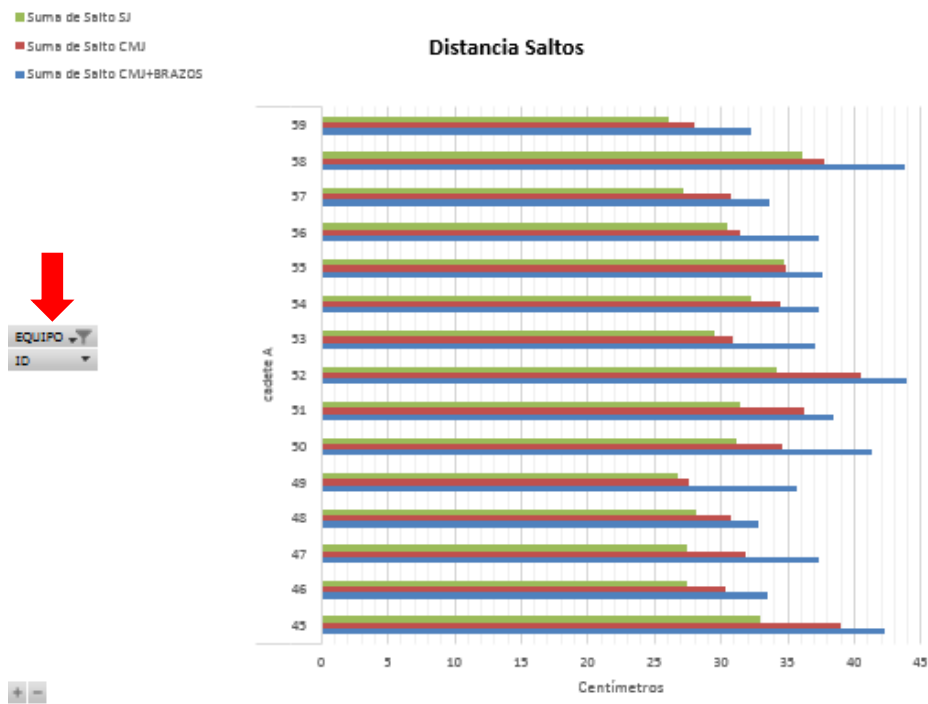
**Figura 55.** Grafica saltos por posición.



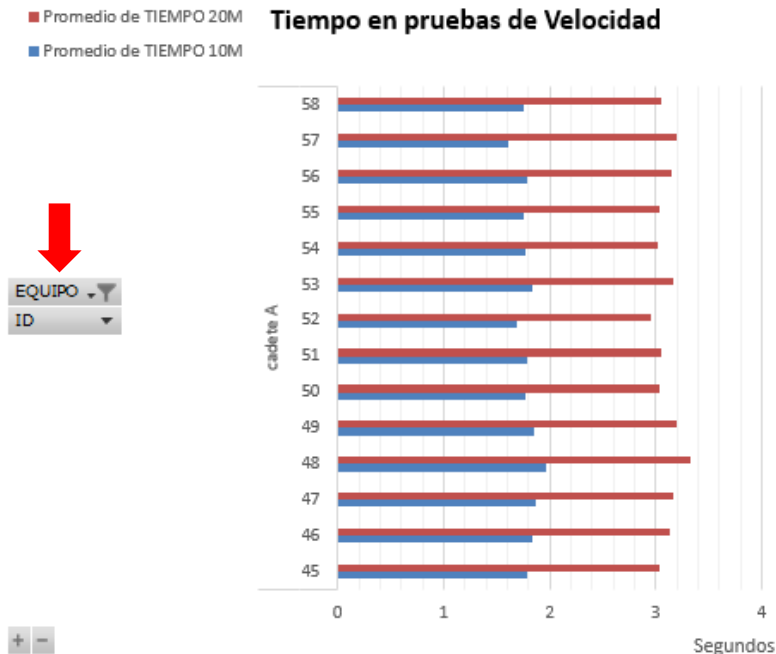
**Figura 56.** Grafica velocidad por posición.



Porwer Pivot tiene gran variedad de herramientas que permite de forma automática modificar los gráficos en función de lo que se desea mostrar, por ejemplo, en la **Figura 57** y la **Figura 58** se representan los resultados de las pruebas de salto y de velocidad para todos los jugadores de un equipo, en este caso el Cadete A. Seleccionando el botón equipo (señalado con una flecha roja), se puede seleccionar el equipo del que se desea mostrar los datos, sin necesidad de crear otro nuevo gráfico.

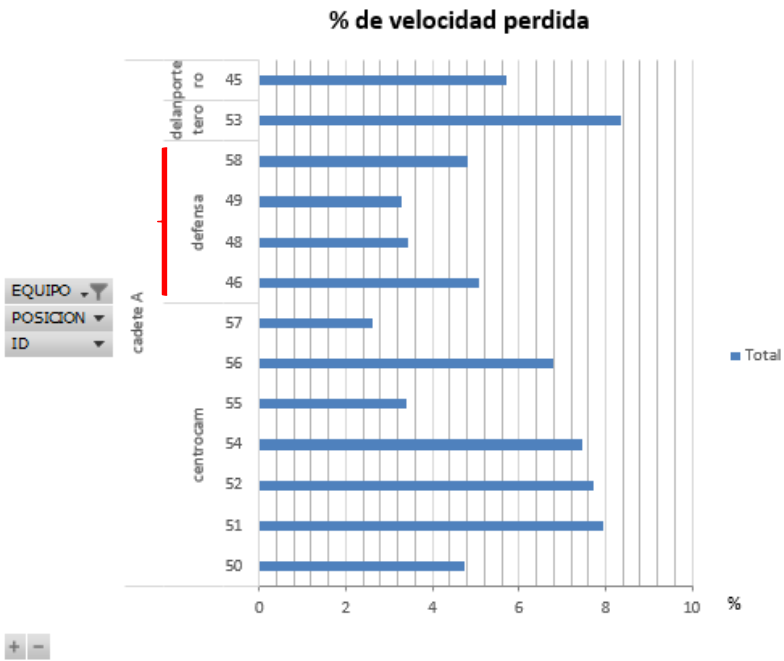


**Figura 57.** Grafica saltos jugadores del equipo Cadete A.



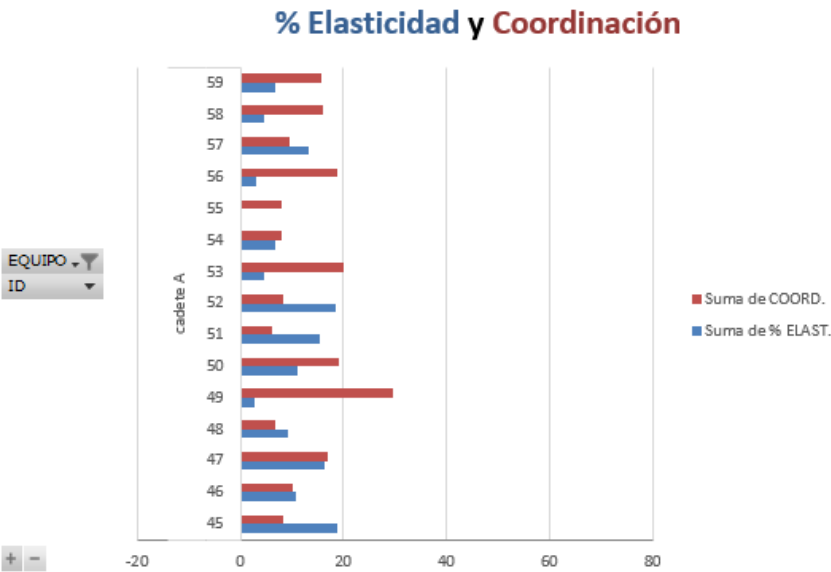
**Figura 58.** Grafica velocidad del equipo Cadete A.

En la **Figura 59** representa el porcentaje de velocidad perdida cuando se hacen 6 series de 20 metros cada una. Se puede observar una característica de los defensas del equipo cadete A, estos obtienen poco porcentaje de perdida de velocidad. Al hacer series de un ejercicio mantienen su velocidad, no necesariamente tienen que ser los más rápidos, pero recorriendo distancias más grandes tiene la habilidad de controlar su velocidad para llevar siempre el mismo ritmo.



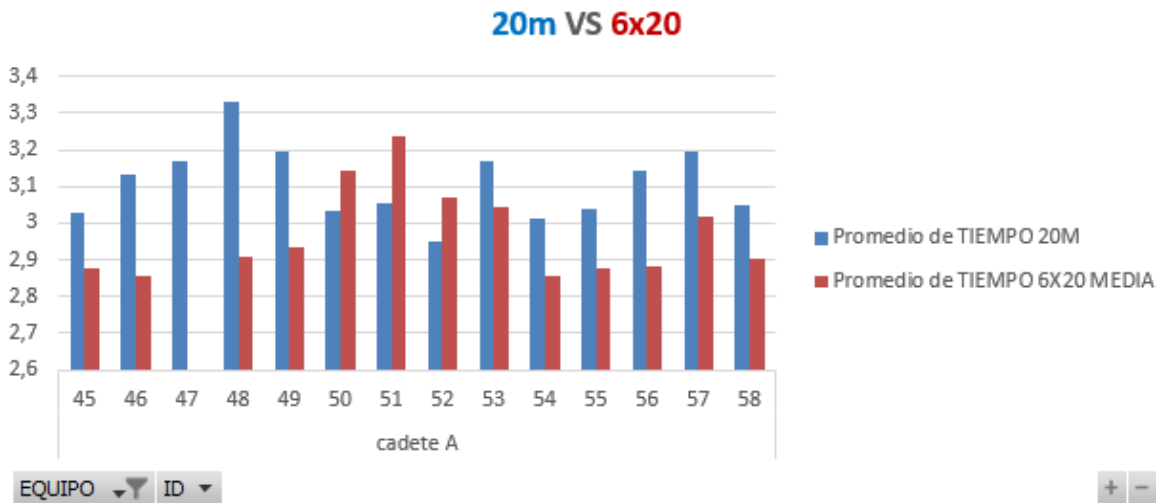
**Figura 59.** Grafica % de velocidad perdida del equipo Cadete A.

La **Figura 60** representa el porcentaje de elasticidad y la coordinación de cada jugador Cadete A. No son datos necesariamente correlativos. La elasticidad debe estar por encima del 10% y la coordinación del 12% para ser clasificado como un resultado favorable.



**Figura 60.** Grafica elasticidad y coordinación del equipo Cadete A.

En la **Figura 61** es representada una comparativa entre el tiempo tardado en hacer 20 metros y la media del tiempo tardado en hacer 6 series de 20 metros. Esto permite visualizar que jugadores tienen mejor rendimiento pruebas de resistencia o de velocidad explosiva.



**Figura 61.** Grafica velocidad 20metros frente a 6x20 metros del equipo Cadete A.

Todas las gráficas descritas con anterioridad se resumirán en un cuadro de mando como el de la **Figura 62**, de este modo con un golpe de vista se podrá visualizar toda la información representada gráficamente.



**Figura 62.** Representación gráfica de un cuadro de mando hecho con Power Pivot.

Estos gráficos que componen el cuadro de mando son gráficos dinámicos, el usuario podrá rediseñarlos en función a sus necesidades de una manera muy sencilla, y a la hora de modificar los datos o agregar más, los gráficos se actualizarán de forma automática. Se podrán realizar informes de los distintos equipos que forman el club sin necesidad de crear nuevos gráficos, simplemente cambiando los valores de los filtros de esto.

## 4. Indicadores de rendimiento (KPIs)

---

A raíz de los análisis realizados hemos podido identificar con éxito diversos factores que resultan clave para el desempeño de los jugadores. Uno de los factores más influyentes es la edad, especialmente debido a que los jugadores se encuentran en pleno desarrollo corporal, lo cual afecta a su fuerza física en gran medida. Por otra parte, y como es lógico, el peso es un factor altamente ligado a la edad, puesto que conforme se van desarrollando el peso va aumentando, lo que cabe la pena destacar, es el descubrimiento de una relación en cuanto al peso total de los jugadores y el peso de su musculatura, aumentando estos dos de manera proporcional, mientras que la grasa no aumenta a medida que aumenta el peso de los jugadores. Por lo que se puede decir que son sujetos con gran masa muscular ya que de su peso depende mucho el músculo.

Debido a esta relación entre la edad y el peso, los jugadores de mayor peso son los que más saltan, y mejores resultados obtienen en pruebas de velocidad y agilidad. Mientras que los de menor peso son más elásticos. Por lo que se debe aumentar los ejercicios de elasticidad conforme vayan creciendo los jugadores, ya que con el paso de los años y el aumento de su peso se va perdiendo esta habilidad.

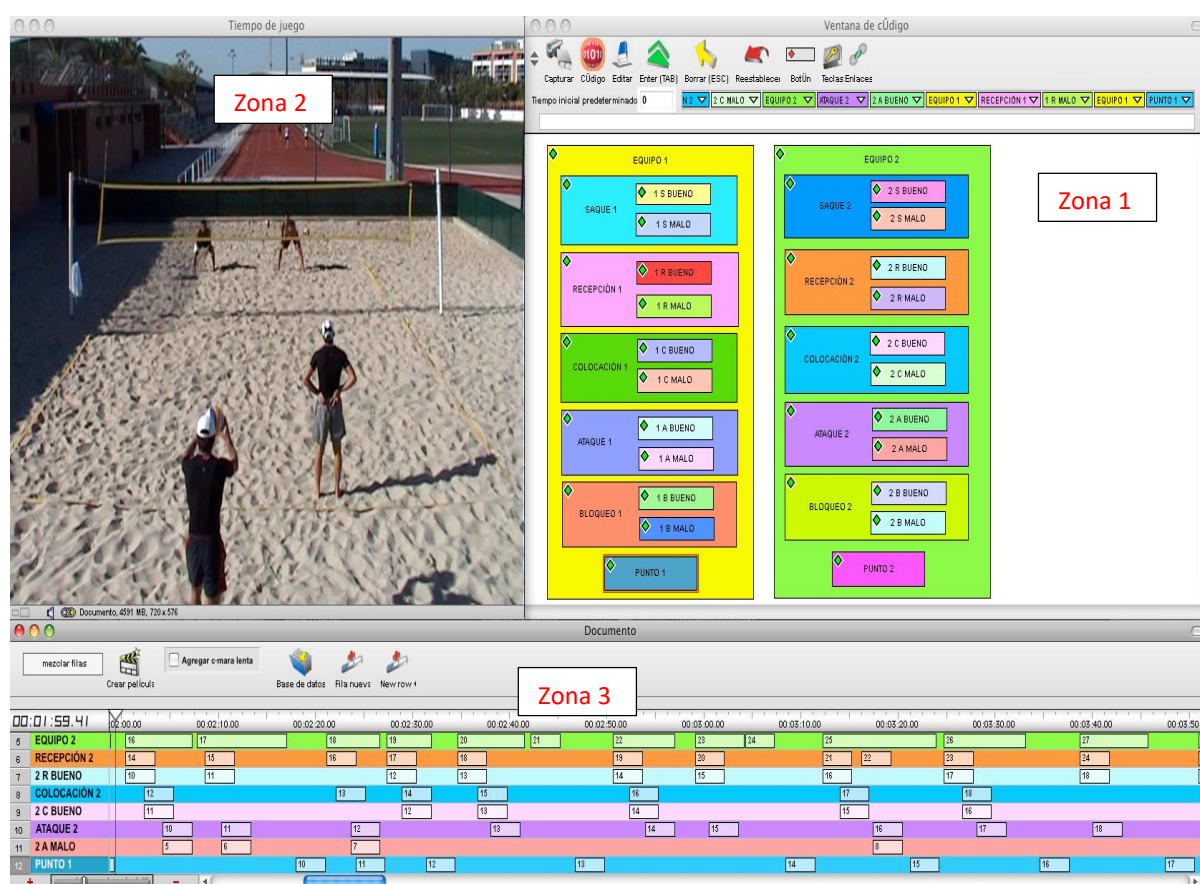
Por otra parte, la posición es un factor indirecto determinante en la velocidad y salto debido a que influye de forma crítica en su rutina habitual y, por tanto, acaba afectando a los resultados del jugador. Por ejemplo, los porteros son los que peores resultados tienen en las pruebas de salto y velocidad, pero obtienen la mayor flexibilidad. Cada posición es determinada por las siguientes características: coordinación y flexibilidad en los porteros, la potencia de salto y velocidad en los centrocampistas y la elasticidad en los defensas.

Se puede observar una característica de los defensas del equipo cadete A (los de mayor edad por debajo del primer equipo), estos obtienen poco porcentaje de pérdida de velocidad. Al hacer series de un ejercicio mantienen su velocidad, no necesariamente tienen que ser los más rápidos, pero recorriendo distancias más grandes tiene la habilidad de controlar su velocidad para llevar siempre el mismo ritmo.

Con toda esta información, podemos definir una serie de indicadores clave de rendimiento para determinar el rendimiento de un jugador teniendo en cuenta su contexto. Para la evaluación de los indicadores tendremos en cuenta la posición y el equipo en el que juega el jugador, dos elementos clave para interpretar los resultados mostrados. Como indicadores, utilizaremos la información de salto, velocidad, flexibilidad y agilidad, que nos permitirán comparar los resultados obtenidos por los jugadores, siempre teniendo en cuenta la posición y el equipo, para los cuales se han definido distintos criterios de éxito como hemos visto anteriormente. Esto permitirá a los entrenadores identificar si un jugador concreto necesita atención especial, ya que resulta deficiente en una o varias áreas comparado con otros en su posición, o si puede ser conveniente cambiarlo de posición debido a sus aptitudes físicas, obteniendo así un mejor resultado para el equipo.

## 5. Perspectivas futuras

Fijando la vista en el futuro, y con la intención de a expansión del proyecto, contamos la colaboración de la empresa Kinetic Performance quien dispone de un software llamado SportsCode (**Figura 63**), el cual analiza de una forma muy eficaz videos deportivos. Este programa permite registrar, contabilizar y producir datos de una forma rápida y eficaz mediante la observación de videos de juego.



**Figura 63.** Captura del programa SportCode.

Esta es una captura del software. Antes de comenzar se personalizan distintos botones en la zona 1, repartidos de forma que resulte más cómodo para el análisis. Una vez se hayan definido todos los botones se reproduce el video de la zona 2, y mediante el pulsado de los distintos botones se crea en la zona 3 cortes de video de las distintas características medidas. El objetivo de dicho análisis es variado, puede ser para conocer en patrón de juego (zonas o acciones más frecuentes), reconocer errores, extraer estadísticas de eficiencia, tener información del oponente real y actualizada, etc. Una vez realizado el

análisis del video este programa genera un documento Excel con los datos que se han obtenido.

Nuestra colaboración consistirá en obtener datos de juego con la herramienta descrita y proceder a su análisis y extracción de resultados como hemos hecho en la realización de este proyecto. Sera construido un puente entre los datos obtenidos por la aplicación y nuestro software, de manera que, tras analizar los videos, automáticamente la información se cargue y permita automáticamente su consulta con los datos actualizados.



## 6. Conclusiones

---

En este trabajo se ha llevado a cabo la aplicación de business intelligence en el ámbito deportivo, una forma de demostrar la potencia de estos sistemas, así como el desarrollo de un sector que no está tan explotado como se podría, ya que pocas entidades deportivas cuentan con esta tecnología.

El objetivo fundamental ha sido la creación de una serie de herramientas para mejorar el entrenamiento de los distintos jugadores en función a los resultados obtenidos en una serie de pruebas físicas, así como facilitar la visualización y el seguimiento de los resultados de dichas pruebas. Para esto se han identificado los factores más importantes que afectan a su rendimiento mediante el estudio de distintas técnicas de Inteligencia de Negocio y Minería de Datos.

Gracias al sistema desarrollado y a los análisis llevados a cabo, se han podido determinar los factores más importantes. En primer lugar, la edad supone un factor determinante en la obtención de los resultados, al realizarse el estudio en jugadores con una edad comprendida desde los 8 años, hasta jugadores pertenecientes al primer equipo. Los jugadores se encuentran en pleno desarrollo corporal, lo cual afecta a su fuerza física en gran medida. Asimismo, la edad está ligada al peso, por lo que se puede extraer que, a una mayor edad del jugador, más masa muscular tiene y mejores resultados obtiene pruebas de salto, velocidad y agilidad. Para la obtención resultados menos influenciados por la edad sería interesante la realización de evaluaciones centradas en grupos nutridos de jugadores de la misma edad, de manera que estos factores no alteren los resultados y puedan ser observados el resto de factores de la muestra con mayor claridad.

Relacionado también con el peso y la edad, se ha observado su relación con la elasticidad. Con el paso de los años y el aumento de peso pierde dicha habilidad, por lo que se deben aumentar los ejercicios de elasticidad conforme vayan creciendo los jugadores.

Por otra parte, la posición también ha sido un factor determinante en las pruebas de velocidad y salto debido a la influencia de su rutina habitual que acaba afectando a los resultados de dichas pruebas.

El equipo al que pertenece el jugador (que viene determinado por su edad) y su posición han sido dos elementos claves para la interpretación de los resultados. Se han determinado como indicadores la información de salto, velocidad, flexibilidad y agilidad.

Gracias a este trabajo, los entrenadores pueden contar con un sistema diseñado exclusivamente para apoyar la toma de decisiones de cara al entrenamiento de los jugadores, mejorando el rendimiento de los equipos y potenciando las características necesarias de cada jugador para cada posición, así como la disposición de un cuadro de mando en el que controlar todos los datos y poder realizar un seguimiento más detallado de la evolución de los jugadores o equipos de manera que de un golpe de vista se puedan comparar todos los resultados y extraer conclusiones.

En el futuro se plantea la realización de dicho análisis enfocado a individuos de la misma edad para obtener resultados menos influenciados, así como la colaboración de la empresa Kinetic Performance y su software SportsCode del cual extraer información a incluir automáticamente en el sistema y poder llevar a cabo su análisis.

## 7. Bibliografía

---

- Business intelligence:

[http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)

<http://www.gestiopolis.com/business-intelligence-teoria-y-conceptos/>

<http://www.gestiopolis.com/inteligencia-de-negocios-business-intelligence/>

<http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/11-introduccion>

- Sistemas OLAP:

[http://www.sinnexus.com/business\\_intelligence/olap\\_vs\\_oltp.aspx](http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx)

<https://jorsodom.wordpress.com/2013/09/19/cual-es-la-diferencia-entre-una-base-de-datos-oltp-y-olap/>

<http://blog.classora.com/2013/06/25/bases-de-datos-multidimensionales-olap-vs-oltp/>

<http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager#x1-540003.4.7>

- Almacenes de datos:

<http://www.dataprix.com/datawarehouse-manager#x1-510003.4.5.2>

- Herramientas ETL:

[http://etl-tools.info/es/bi/proceso\\_etl.htm](http://etl-tools.info/es/bi/proceso_etl.htm)

<http://www.dataprix.com/33-load-manager>

- Herramientas de creación de informes:

Pentaho Report Designer:

<http://wiki.pentaho.com/display/Reporting/Pentaho+Reporting+Community+Documentation>

Power BI: <https://powerbi.microsoft.com/es-es/>

Power pivot: <https://support.office.com/es-es/article/Power-Pivot-an%C3%A1lisis-de-datos-eficaz-y-modelado-de-datos-en-Excel-d7b119ed-1b3b-4f23-b634-445ab141b59b>

- Desarrollo del sistema

<http://mondrian.pentaho.com/documentation/workbench.php>

<http://community.pentaho.com/projects/mondrian/>

<http://wiki.pentaho.com/pages/viewpage.action?pageId=14844841>

- Análisis

Weka:

<http://www.cs.waikato.ac.nz/ml/weka/>

<http://users.dsic.upv.es/~cferri/weka/>

Power pivot: <https://support.office.com/es-es/article/Tutorial-Crear-informes-de-Power-View-sorprendentes-Parte-1-e2842c8f-585f-4a07-bcbd-5bf8ff2243a7?ui=es-ES&rs=es-ES&ad=ES>